THE UNIVERSITY OF ALBERTA

ASSESSING INTELLECTUAL ABILITY WITH A MINIMUM OF CULTURAL BIAS FOR TWO SAMPLES OF METIS AND INDIAN CHILDREN

by

Lloyd Wilbert West

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF MASTER OF EDUCATION

DIVISION OF EDUCATIONAL PSYCHOLOGY

Edmonton, Alberta

Date: July, 1962.

ABSTRACT

The purpose of this investigation was to ascertain the effects of cultural background on selected intelligence tests with a view to identifying instruments which, in a cross-culture administration, validly measure intellectual ability with a minimum of cultural bias.

A battery of tests, selected by Dr. R.S. MacArthur for a factor analytic study was administered at four grade levels to a sample of 126 Metis and Indian children attending the Faust School. As a replication, an identical battery of tests was administered at the same grade levels to a sample of 155 Indian children attending the Fort Simpson School.

The relative extent of cultural bias in the tests for each sample and at each grade level was determined by comparing the mean of the derived scores on the tests. All such derived scores were based on a scale of Edmonton or Calgary T-scores in order to assure comparability of performance on the various tests and to permit a comparison of performance between cultures.

The batteries of tests employed in this study generally included conventional intelligence tests and subtests as well as several promising non-verbal or culture-reduced varieties.

The major findings for the study were as follows:

1. Some tests show significantly less cultural bias than do others. The concept of cultural bias in tests, therefore, appears to be well founded.

2. Tests which:

(a) consist largely of items that can be solved in any language or mode of expression.

(b) have minimal dependence on past specific learning, and

(c) are probably as novel to one culture group as to another,

show significantly less culture bias than other more conventional tests, and therefore may be appropriately termed "culture-reduced".

3. These culture-reduced tests show significantly less increase with grade level as a result of selection and school treatment than do traditional educationally loaded tests.

4. Culture-reduced tests show substantial correlations with academic achievement and therefore must sample from the abilities required for academic success.

5. A test which appears to show little cultural bias at one level may show considerably greater bias at another level and vice versa.

The tests of the experimental batteries used in this investigation were evaluated against four criteria. Those tests which show greatest promise, at each level for a cross-culture assessment of intellectual ability have been identified.

ACKNOWLEDGEMENTS

Several persons provided generous assistance in the research project upon which this thesis is based. The writer wishes to express sincere thanks first to his supervisor Dr. R.S. MacArthur whose continual assistance was well beyond the call of duty and who provided the greater part of the data for the project; to Dr. W.B. Elley for his helpful comments and thoughtful criticisms and whose Ph. D. thesis was a source of invaluable guidance; to Dr. S.M. Hunka for his supervision of the organization of data, card punching, and electronic computing.

Many thanks are also due to Mrs. J.I. West and Mrs. E.P. West for their many hours of valuable clerical assistance; to Mrs. H.M. Trahan for her careful and prompt secretarial work; to Miss. Greta Tratt for her prompt and efficient operation of the I.B.M. 1620 in the processing of much of the data.

The writer is indebted also to the Province of Alberta for a Province of Alberta Graduate Scholarship and to the Alberta Teachers' Association for the John Walker Barnet Scholarship, which together provided liberal financial assistance.

It is with sincere gratitude that the contribution of each is acknowledged.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

THE GENERAL PROBLEM

Intelligence and Cultural Differences

It is generally agreed that the performance of an individual on an intelligence test is to some extent a function of the kinds of experience he has had. Hebb (1958) cites studies of children growing up on canal boats in England, and in isolated mountain communities in United States as supporting evidence that "cultural differences affect intelligence test scores to a very marked degree." Elley (1961) cites several other studies which lend support to the same hypothesis. Eells (1951) states that "the existence of these differences must today be accepted as a demonstrated fact."

To the extent that a measure of intelligence is a function of cultural factors and therefore discriminates against members of various cultures it may be said to be culturally biased. A series of studies is being conducted at the University of Alberta, under the supervision of Dr. R.S. MacArthur, in the general area of assessing intellectual ability with a minimum of cultural bias. Findings thus far, noteably Elley's (1961), corroborate previous evidence that conventional tests of mental ability discriminate against children from cultures other than the

urban, middle-class American. Accordingly, in this investigation, the dependency of measures of intellectual ability upon cultural factors is regarded as a given.

Minimizing Cultural Bias

MacArthur (1961) draws attention to the fact that Canadian philosophies of education expect that instruction will be adapted to the intellectual potential of the pupil. Teaching treatments adapted to pupil potential, however, pre-suppose a fair estimate of that potential. Toward this goal, the search for measures of intelligence which are relatively independent of specific environmental experiences becomes especially important.

Several so-called culture-reduced tests have been developed in an attempt to minimize the effects of cultural differences. Examples of such tests would include the IPAT (Cattell) Culture-Free Intelligence Test, Raven's Progressive Matrices, the Safran Culture Reduced Intelligence Test, and other non-language tests. Anastasi (1961) expresses the concern that such tests may sacrifice predictive validity without appreciably reducing cultural bias. Concern such as this emphasizes the need for further research. A distinction may be made between predicting to specific criteria, given a fixed treatment, and predicting to general criteria, given an adaptive treatment. Although such a distinction, as elaborated by Cronbach and

Gleser (1957) can, in part, answer Anastasi's criticism of the culture-reduced intelligence tests, it would be unwise to give extensive application of such tests in advance of adequate research. We must first know answers to such questions as: "What do the tests measure?" "To what extent do they measure what we want them to measure?" "How reliable are they?" and "To what extent are they free of cultural bias?" The last question is of central interest in this investigation.

The general purpose of this investigation, as part of the Alberta series of studies previously mentioned, is to study a selection of promising culture-reduced tests, and to identify those which for a sample of Faust, Alberta Metis children and another sample of Fort Simpson, N.W.T. Metis and Indian children, show a minimum of cultural bias.

CHAPTER II

THEORETICAL FOUNDATIONS

Introduction.

Although psychological theory has made enormous advances in the past century it is still in its infancy. It has a long way to go before it can explain the principles of behavior, including that variety which we call "intelligent", to the same degree that the physical sciences can explain the principles of physics and chemistry.

The present chapter provides a theoretical framework for a study of the effects of culture on intelligence. Some of the working assumptions are supported by a large body of empirical evidence. Others are more novel and, as yet, have not been submitted to rigorous verification. It is taken for granted that any such theory must, at present, necessarily be inadequate and incomplete.

A theory which becomes static is also likely to become dogma. The creative production of novel hypotheses is necessary for the continued growth and development of any discipline. One must never assume that because a hypothesis has not been verified beyond reasonable doubt that no one in the future will be able to do so. Accordingly in the present quest for psychological knowledge the "merely plausible" must be given consideration. It is quite

likely that many of the working assumptions today will have to be rejected at some later date; but such is the history of all science.

The Concept of Intelligence.

The psychological controversy of "nature" versus "nurture" has not been resolved. Investigators have provided experimental evidence in support of the dominance of each variable. A perusal of the literature leads one to conclude that when environment is fairly homogeneous most of the variability of intelligence may be determined by heredity. On the other hand, when hereditary variation is at a minimum, as among siblings, most of the variability of intelligence is determined by environment. Hence, in this investigation, present level of intellectual functioning will be assumed to be the joint product of heredity and environment. Hebb (1958) supports this position and states that "both these variables are of 100 per cent importance: their relation is not additive but multiplicative". MacArthur (1961) expresses the same theoretical concept when he states that differences in intellectual abilities "result from interaction of innate predispositions and environment in a multiplicative rather than additive manner".

Much of the controversy and confusion regarding intelligence has resulted from different uses of the term.

In order to clarify its useage, Hebb (1949) distinguishes between "intelligence A" and "intelligence B". Intelligence A is an innate intellectual potential, whereas intelligence B is present level of performance. Recognizing that different tests will assign different values to an individual's present level of performance, Vernon (1960) added "intelligence C". Intelligence C is an estimate of B made by an intelligence test.

Behavior can be observed directly, and the adjective "intelligent" can be used to describe certain aspects of its quality. Intelligence has been defined as ability to learn, capacity to adapt, ability to educe correlates, and the like. Wechsler (1958) points out that all these definitions assume "varieties of behavior which might reasonably be called 'intelligent'". The noun "intelligence", however, is not a real observable or measureable entity. It has postulated existence only, like the "self". Its usefulness derives from its ability to unify and explain otherwise isolated empirical facts regarding behavior. Intelligence tests attempt to measure present levels of intellectual functioning or intelligence B. The measure of B for a given individual on a given test at a given time is a function of the number of standardized tasks for which the quality of his response is judged to be intelligent.

There are a great many kinds of intellectual abilities. Thurstone proposed about a dozen group factors which

he designated as "primary mental abilities". He assumed that just as all colors may be composed from the primary colors, so too are all mental abilities a composition of various proportions of the primary abilities. Guilford (1959) postulates as many as 120 or more intellectual factors in his schema which he calls "The Structure of Intellect". The development and use of statistical techniques such as factor analysis has made it clear that many different intellectual factors exist and that intelligence tests sample differentially among these factors. Consequently, the various tests yield measures of somewhat different aspects of intellectual behavior.

Vernon (1956) conceives of intellectual ability from a statistical viewpoint as hierarchial in structure. General intellectual ability similar to Spearman's g is at the top of the hierarchy. It enters into or is extensively sampled by most items of general ability tests. Further down the hierarchy are two broad group factors: verbal-educational (v:ed) and practical-mechanical (k:m). These major group factors can be further analysed or subdivided into minor group factors. The minor group factors in turn, may be subdivided into more specific factors and so on, to any degree of specificity.

The series of studies being conducted at the University of Alberta is in general based upon such a hierarchial concept of intelligence.

The statistical construct g has certain interesting properties which should be noted. Elley (1961) cites studies of Woodrow, Cattell, and others in support of the view that g is relatively immune to environmental influence. He later tersely summarizes the virtues of g as follows:

> It is exemplified in behavior of a complex rather than a simple kind, is relatively stable and independent of differences or changes in environment, and is of considerable value in predicting to non-specific criteria. (p. 52)

On the b asis of previous research and existing literature, it is postulated that g:

(a) is a general ability factor likely to enter heavily into a wide variety of future intellectual performance,

(b) is relatively immune to specific cultural influences,

and therefore:

(c) may be interpreted as a measure of intellectual potential.

Distinguishing Between Proficiency and Potential.

Ability tests vary in construction and use. Anastasi (1961) states that "all ability tests fall along a continuum with regard to their dependence upon specified prior experience". In Cronbach's (1960) schemata, ability tests are arranged in a "spectrum" varying from those of maximum educational loading to those of minimum educational loading. Tests at the various positions of the continuum or spectrum have different uses and different merits. Cronbach (1960) states that those with maximum educational

loading "are designed for cold blooded prediction of future school success", especially where treatments are fixed. The efficiency of these tests for this purpose is given support by evidence summarized by Clarke (1958). He found the median correlation between various measures and university marks to be as follows:

(a)	high school average	r = .56
(b)	achievement tests	r = .49
(c)	scholastic ability tests	r = .47

It will be noted that measures having the greatest educational loading are the best predictors of scholastic success. This is especially true when treatments in the past have been comparable and when future treatments are fixed. Cronbach's further discussion of the functions of tests with minimum educational loading is noteworthy:

> ... the teacher and the counsellor working with a student wants to know what undeveloped resources he has. They can rely on past achievement for an estimate of probable future accomplishment when nothing out-of-the-ordinary is done for the student, but the mental test ought to locate undeveloped potential that novel treatment may bring out. For the latter purpose the most important information is provided by tests ... which have a minimum overlap with achievement. (page 236).

Elley (1961) suggests that in an attempt to improve their predictive ability, tests which purport to measure intelligent behavior have been moved farther and farther into the realm of achievement and past experience. The difference between measures of proficiency and measures of potential is largely a matter of the extent to which the tests sample specific past learning. Conventional

intelligence tests characteristically use the same kinds of problems that are found in achievement tests. They are also usually validated in terms of school achievement. The correlation between measures of intellectual potential and school achievement should be only moderate. Many factors, other than intellectual potential, such as effort, teacher bias, previous school background, and physical and emotional health affect school achievement. The need for ability tests with minimum educational loading is therefore clear.

Educationally loaded tests measure information quite as much as thinking power. Good (1954) points out that we "cannot measure ability directly; what we measure is performance ... from which we infer ability, from which we infer capacity". In agreement with this point it is suggested that:

(a) all ability tests measure performance,

(b) from this measure we may only infer intelligence B,

(c) under certain conditions, from intelligence B, we may infer intellectual potential.

The conditions under which the inference of potential may be made is illustrated with reference to Anastasi's (1961) p. 486, continuum of ability tests.

Maximum educational loading		Minimum educational loading	
———————————————			All tests measure performance.
	———————————		From these we may infer intelligence B.
		————————	From these we may, with caution, infer potential.
Assessing proficiency in special areas	Assessing general intellectual ability	Assessing intellectual potential	) Major Use.

As we move toward minimum educational loading we may make inferences about both present level of intellectual functioning and potential for future intellectual development with increasing confidence.

In addition to their importance for assessing the potential of children from culturally handicapped backgrounds, tests with minimum educational loading (ie. the culture-reduced tests) are of great value in the identification of under-achievement. This task is tantamount to the location of appreciable deviation between present level of performance and intellectual potential. Thus,to the extent that estimates of achievement and estimates of capacity overlap or measure the same thing they are ineffective in the identification of under-achievement.

Source of Differences in Measures of Intelligence.

Elley (1961) has reviewed and summarized professional opinion on the causes of differences in measures of intelligence between different cultural groups. From the more prominent interpretations in the literature it may be concluded that three main variables are regularly used to account for the variance in measures of intelligence B:

(a) Hereditary differences

(b) Differences in opportunity to develop intelligence

(c) Cultural bias of the test.

i. Item dependency on information or skills that not all children have had equal opportunity to acquire.

ii. Dependence upon acquired personality factors more likely to be developed in one culture than another, eg. motivation, work habits.

Many psychologists emphasize the importance of one variable to the exclusion or neglect of others. In this study multiple causation is assumed. Each of the three variables listed is regarded as a significant contributor to the differences in measures of intelligence. The importance of each varies according to the extent to which the other variables are held constant.

The Development of Intelligence and the Critical Period Hypothesis.

Intelligence A is the innate potential for those organized patterns of behavior which we describe as intelligent. It is but the substratum upon which intelligent behavior is developed. This "fully innate property", Hebb (1949) suggests, "amounts to the possession of a good brain and a good neural metabolism". Beginning with these biological "raw materials", the acquisition of mental ability or intelligence B, is the result of developmental processes.

According to Piaget (1950) none of our mental abilities are innate, and none develop in isolation. By exploration of the environment every concept is gradually built. The sensory-motor and perceptual associations formed in early life are basic to the later images and ideas which we symbolize in language and thought.

Hebb's (1949) work in animal psychology has led him to similar conclusions. According to his neuro-psychological theory, early experiences result in permanent changes in the organization of pathways in the cerebrum. Environmental stimulation results first in the establishment of cell assemblies and secondly in an interfacilitation between these pathways. All learning tends to utilize and build upon earlier learning instead of replacing it.

The learning of the mature animal is thus based upon, limited by, and canalized by the slow inefficient learning that has gone before.

There is a great deal of evidence that environmental deprivation and depressing emotional climate restrict the development of intelligence, (Lynn, 1959; Goldfarb, 1955). Vernon[1] lists several factors which may inhibit intellectual development.

- poverty
- ill health
- parental attitudes which demand that the child be passive, submissive, inhibited, and obedient
- unstable family life
- lack of a stable father figure with whom to identify
- schooling which is abbreviated and irregular
- formal authoritative and punitive discipline
- language handicaps

Both Bayley (1957) and Wechsler (1958) found that different intellectual functions develop at different rates with different ages contributing maximally to their total development. The growth curves of such studies suggest that the development of an intellectual function is most efficient at a certain stage of total development. This observation leads to a very reasonable question: What will be the end result of a curtailment in development during that period in which maximum development is normally expected? May it not well be that certain periods are critical for the development of an intellectual ability?

---

[1]Lecture delivered at the University of Alberta. April, 1962.

The critical period hypothesis holds that, in the development of an organism certain processes or functions must occur in a certain sequence and within a given time limit or they may never occur, but instead their potential may atrophy. While in the process of growth, a function or organ is much more susceptible to change produced by environmental agents than it is when fully formed. The critical period hypothesis holds beyond doubt in many non-intellectual areas of development. For example, in the area of physical development, it is not difficult to influence height at an early age by nutritional or hormonal variation; but it is virtually impossible to affect it once the individual has completed his development.

As early as 1875, Spalding found that the flying ability of swallows could be permanently impaired by restricting their incipient flight movements from an early age. Dennis (1941) repeated the experiment on another species of birds and obtained even more conclusive results. Riesen (1950) was able to show that a critical period exists in the development of sight in chimps. Severe and enduring deficits in intelligent behavior of rats and dogs as a result of early restriction have been shown by a large number of workers. (Hebb, 1947; Forgays and Forgays 1952; Hymovitch 1952; Bigham and Griffiths 1952; Thompson and Heron, 1954; Clark et al. 1951).

The work of both Piaget and Hebb may be interpreted

as lending support to the applicability of the critical period hypothesis to the development of human intelligence. Both suggest that it is during the early years of life, while a large part of the brain is still developing, that the basis for all complex psychological processes of perceiving, learning, and remembering are laid down. Hebb (1958) states that the "effects of stimulation during the growth period are in great degree permanent for good or ill so that the level of intelligence B which is achieved by the age of fifteen or thereabouts will last".

Bayley's (1957) work in the Berkley Growth Study has led her to postulate the possibility of "environmental 'critical stages' for mental growth". After reviewing the evidence regarding the effects of early environment on later behavior, Thompson (1955) concludes as follows:

> There is no doubt that an individual continues to learn and perceive throughout his life span; but it is probable that he learns to learn and perceive quite early, and that once these basic capacities are fully developed they cannot be changed very much. It also seems true that there are certain periods in early life when a particular function or structure is maximally susceptible to change. It is probable that the length of these critical periods as well as their time of occurrence, varies considerably with different functions and structures.

Thompson has also developed a neurological explanation for the critical period hypothesis. While this explanation goes far beyond the evidence available at present it certainly appears to be one worth exploring.

If it is assumed that intelligence tests measure all

round ability in thinking skills and that such thinking capacity is based upon experience and appropriate stimulation, the principle of selective attention becomes relevant. A learned predisposition to engage in certain experiences or to discriminate, note, and attend to certain environmental stimuli may facilitate intellectual development. It is therefore possible that the individual "selects" the abilities he will develop according to the criteria of compatibility with his basic personality structure. A study by Haggard (1957) suggests this possibility. It is well known that individuals in some cultures develop a high duplicity quotient although they may fail to develop a correspondingly high intelligence quotient. Children from cultures other than urban, middle-class, American may not avail themselves of the appropriate stimulation which develops the symbolic skills through which we measure intelligence simply because such skills are not significant in the style of life which they have defined for themselves, or which their culture has defined for them. Anastasi (1958) has noted that cultural and racial differences in intelligence scores may result from the values stressed and the goals set for growth in a given culture. Children in cultures which stress physical skill may give little attention to intellectual abstractions. Those from cultures that stress intellectual attainments, on the other hand, may be motivated to do well in them. Appropriate attitudes

may therefore be quite as important as the availability of appropriate experiences for the development of intellectual abilities.

Although certain critical periods early in life may contribute maximally to the development of an intellectual function there is considerable evidence that the structure and level of intellectual abilities may to some extent be modified during the adolescent years. According to Garrett's (1946) developmental theory of intelligence, the dominance of g in intellectual performance decreases with age and specific factors emerge as a result of increasing differentiation of ability. The findings of Meyer and Bendig (1961) suggest that each of Thurstone's PMA tests measure essentially g for young children. Only in late adolescence did they find the emergence of specific abilities which, incidentally, they attribute to training. Bradway and Robinson (1961) found that intellectual growth continues well past adolescence, and that whatever the nature of the relationship between intelligence of children and their parents, the main effects have occurred by the time the child reaches adolescence. They found that although the correlation between I.Q. of parent and child increases with age up to adolescence, at this point it begins to decrease. Vernon (1958) found that after allowing for individual differences on the 11+ examinations, children who attended the best secondary schools gained 12

I.Q. points over those who attend the weakest schools. Husén (1951) in Sweden and Lorge (1945) in America have found that adult I.Q. scores are affected by the amount of schooling received. These changes probably represent learned increments in the specific skills sampled by conventional intelligence tests. By piling up educational experiences, the better schooled, obviously can be expected to increase their scores on educationally loaded tests. General intellectual ability (g) measured by culture-reduced tests may be less affected by such later learning opportunities.

Present Potential; Summary, and Conclusions.

From the assumption that intelligence B results from the interaction of innate predispositions and environmental influences in a multiplicative manner certain interesting logicomathematical deductions may be made. First, the assumption may be expressed by the mathematical model:

B = AE where:

B is Hebb's intelligence B, present level of intellectual functioning, or that which intelligence tests measure.

A is Hebb's intelligence A, or the innate substratum of predispositions or potentialities upon which intelligent behavior is developed.

E is the totality of cultural or environmental experiences, which stimulate, thwart and direct organized behavioral development.

From this paradigm it becomes evident that what we hope to measure when we attempt to hold E constant is an

estimate of A. The magnitude of A can never be measured directly but can only be inferred from the magnitude of B, yet B is a function of E. By simple algebraic substitution, setting E equal to zero makes B also equal to zero. The implication is that it is impossible to design a test of B which is not contingent upon E. It also follows that psychometrically we can never arrive at an estimate of A which is independent of E.

However impossible it may be to design a test which is "culture-free" we must not abandon the task of designing "culture-reduced" tests. Elley (1961) defines a culture-reduced test at "one which holds verbal symbolism and past learning to a minimum and which uses a medium which is relatively unfamiliar to all examinees."

In developing a rationale for culture-reduced tests, it may be assumed that, since all men live on the same planet, belong to the same animal species, and have the same biological needs, they may share, but with considerable variation, certain universal experiences of crucial importance to the development of intelligent behavior.[1] Our efforts

---

[1]The experiences postulated here are general rather than specific in nature. They are initially encountered during the first few years of life. They include a variety of stimulation such as:

- the perception of motion and other objects in space.
- sounds of varying pitch and amplitude.
- hunger, thirst, pain, and comfort.
- random attempts to make adjustment to the environment.
- satisfaction and frustration of basic needs.
- learning to grasp and manipulate simple objects.
- locomotion (learning to crawl or walk).
- communication (learning to use some form of symbolic representation or language).

to achieve better estimates of A by reducing the cultural effect of the measuring instrument may take the form of identifying test items which sample behavioral organization capable of growing out of the common experience of man. This is what available culture-reduced tests have tried to accomplish. In practice such tests, in many respects, fall short of this objective. No test is entirely free of cultural influence. The difference between culture-reduced and conventional tests is a matter of degree.

Bayley (1940) and Bayley and Jones (1937) found that the cultural background of parents is related to children's intelligence scores after the first two years, but not in infancy. Up to a certain stage in infancy the minimal requirements for life may furnish adequate stimuli for normal intellectual development. It is as children grow older that the varied experiences provided by different cultural environments show marked effect on mental development. Hence increasing differentiation between children of different cultures with increasing age is an indication of cultural bias in the test.

If the concept of "potential" or "capacity" in contradistinction to proficiency is to have any place in psychological theory it is necessary to distinguish between intelligence A (ie. innate potential) and present potential. In order to clarify this distinction, reference willbe made to the following codification of the factors involved in development which was suggested by Hebb (1958) p. 121.

Class of Factor	Source, Mode of Action, etc.
I. Genetic	physiological properties of the fertilized ovum
II. Chemical, prenatal	nutritive or toxic influence in the uterine environment
III. Chemical, postnatal	nutritive or toxic influence: food, oxygen, water, drugs, etc.
IV. Sensory, constant	pre- and postnatal experience normally inevitable for all members of the species.
V. Sensory, variable	experience that varies from one member of the species to another.
VI. Traumatic	physical events tending to destroy cells: an abnormal class of events to which the animal might conceivably never be exposed.

Present potential refers to the individual's present capacity for future development of intelligent behavior. It depends on how well he has sustained the effects of factors II, III, IV, V and VI in his development. If there are critical periods for the development of an intellectual function, factor V becomes extremely important. In this event present potential will also be a function of the individual's success in the various critical periods that have already elapsed. It should be noted, however, that the assumption of the critical period hypothesis is not essential to the concept of present potential. Present potential must

always be equal to or less than intelligence A. It would be equal only if past environmental conditions, both prenatal and postnatal, had been optimum -- a condition which cannot usually be assumed. For convenience we may refer to present potential as intelligence A'. A' is the residual of an individual's original innate potentiality at any time after conception. Whereas intelligence A is a constant, intelligence A' is a variable such as $A' \leqslant A$. The essential difference between intelligence A' and intelligence B at a given time is that A' refers to the potential or capacity for future development given novel and optimum treatment. Intelligence B on the other hand refers merely to the level of presently developed and functioning intellectual abilities. Although analogies are risky in scientific discourse, it may help to clarify the distinction by comparing intelligence A' to the state of the "raw materials" at any given time, and intelligence B to the "developed product" at the corresponding time.

It is suggested that an estimation of the magnitude of intelligence A is even more remote than previously supposed. An evaluation of A would have to be based not only on the assumption that all subjects to be tested had had common experiences sufficient to develop to full potential whatever abilities the tests measure, but also that factors II to VI of Hebb's codification have not intervened to decrease the individual's innate potential.

It was previously suggested that if we could eliminate the variance attributable to E from our measures of intelligence B we would have an accurate estimate of intelligence A. The major difficulty in getting an estimate of A is seen to lie in our inability to eliminate the effects of the developmental factors II to VI. The composite variable E, includes all these factors.

There is one area of environmental differences which may be partially controlled however, and which contributes to what has been called the cultural bias of a test. This area includes,

(a) specific information and skills that not all children have had equal opportunity to acquire, and,

(b) personality factors more likely to be developed in one culture than another, eg. motivation and work habits.

Many promising attempts have been made to minimize the influence of these differences in the design of measuring instruments; eg. the culture-reduced tests.

It is suggested that if we minimize the variance in measures of B which is attributable to these differences, and if we disregard the variance attributable to other environmental factors, the resultant measure will be an estimate not of intelligence A but of what has been referred to as present potential or intelligence A'.

It appears futile at present to attempt even an indirect estimate of intelligence A for cross-culture

comparisons, but the assessment of intelligence A' is a realistic goal. This is a fortunate situation as it is present potential, in contradistinction to innate potential, that is of interest and practical significance to the educator. It is present potential on which future development depends and on which adaptive teaching treatments must be based.

In summary, it is postulated that tests heavily loaded with the statistical factor g and having minimum cultural loading provide our best estimates of present potential.

CHAPTER III

DEFINITIONS, POSTULATES, AND HYPOTHESES

Definitions

1. non-language test - a test which requires no knowledge of reading and writing.
2. culture loading - the extent to which a test is dependent upon types of learning experiences to which members of one culture have a higher probability of exposure than those of another culture.
3. culture-reduced test - a test which has minimal dependence upon verbal symbolism and past learning, and the medium of which is relatively novel to all subjects.
4. cultural bias - operationally defined, in this thesis, as the extent to which a test discriminates between, on the one hand, Metis and Indian children of the Faust and Fort Simpson samples and, on the other hand, urban children of the Edmonton and Calgary samples.
5. cross-culture test - any instrument designed or adapted for testing individuals reared in different cultures including the tests variously designated as non-language, culture-free, culture-fair, or culture-reduced.
6. intelligence A - the innate substratum of predispositions or potentialities upon which intelligent behavior is developed.

7. intelligence B - the present level of intellectual functioning of an individual.
8. intelligence A' - the present potential of an individual for future development of intelligent behavior.
9. E, environmental influences - the totality of cultural or environmental experiences which stimulate, thwart, and direct organized behavioral development. The unique permutation of unit experiences of an individual taken from the infinite variety possible.
10. Metis - a child who, according to his teacher, is considered by the community to have Indian "blood in his veins".

Postulates

1. Individual differences in measured intelligence result from the interaction of heredity and environmental forces in a multiplicative manner.
2. The first postulate may be accurately represented by the mathematical model B = AE, where B is intelligence B, A is intelligence A, and E is environmental influences.
3. Environmental influences encompass cultural differences and affect measures of intelligence B in two major ways:
   (a) by providing differential opportunities to develop intellectual capacities.
   (b) by providing differential opportunities to learn the specific information, skills, attitudes, work habits, etc. required for successful performance on the test.

4. The effect of minimizing the total influence of E on the measure of B is to maximize the influence of A.
5. The effect of minimizing the cultural bias of a test (ie. minimizing the effect of b) in postulate 3) is to maximize the influence of A'.
6. Tests or sub-tests which depend upon verbal symbolism and acquired information differentiate more clearly between urban white children and rural Metis and Indian children than do non-verbal reasoning tests.
7. The statistical factor g or general ability is likely to enter heavily into a wide variety of future intellectual performances and may therefore be interpreted as a measure of present potential.
8. Culture-reduced tests are less dependent on the contingencies of environmental stimulation than conventional tests, and are therefore able to provide better estimates of present potential.
9. The following criteria adapted from those suggested by Elley (1961) should be used to evaluate a cross-culture test.

   (a) it should largely sample a broad general factor of intellectual ability.

   (b) it should show negligible loadings on verbal, numerical, and other group factors.

   (c) it should show less difference between cultures in a bi-cultural administration than do conventional verbal intelligence tests.

(d) it should show relatively little decline with age for children from culturally handicapped backgrounds.

(e) it should contain items that can be solved in any language or mode of expression and which are likely to be as familiar and useful for one cultural group as another.

(f) it should show moderate relationship to school achievement.

Hypotheses

This investigation is exploratory in nature. Its purpose is to provide an answer to the urgent and practical question: "Which of the available psychological group tests or sub-tests, appear to assess the intellectual potential of Metis and Indian children with a minimum of cultural bias?" To accomplish this task various criteria were set up (postulate 9, above) and the following hypotheses were made about the battery of tests administered at different grade levels to samples of Metis and Indian children at Faust, Alberta and Fort Simpson, N.W.T.

1. Some of the tests in the experimental battery administered to a given grade level will show significantly less cultural bias than will others.

2. The tests which show least cultural bias will consist largely of items that can be solved in any language or mode of expression and which are probably as novel to one group as to another. In order to test this hypothesis each of the tests was inspected and categorized into two

groups prior to the analysis of data. Tests which appeared to fit the above description were classified as "culture-reduced". Those which did not were classified as "conventional".

Culture-Reduced Tests	Conventional Tests
Lorge-Thorndike II (Primary)	Lorge-Thorndike I (Primary)
Lorge-Thorndike III (Primary)	Lorge-Thorndike Total (Primary)
Lorge-Thorndike I Non-Verbal (Levels III and IV)	Detroit Beginning First Grade
Lorge-Thorndike II Non-Verbal (Level III and IV)	California Test of Mental Maturity (CTMM) - Total
Lorge-Thorndike III Non-Verb. (Levels III and IV)	CTMM - Logical
	CTMM - Numerical
Raven's Colored Progressive Matrices	CTMM - Verbal
SCRIT (Safran Culturally Reduced Intelligence Test)	
CTMM - Non-Language	
CTMM - Spatial	
Cattell.	

3. The culture-reduced tests common to the experimental batteries for Grade II and III level and Grade VII and VIII level will, for Metis and Indian children, show significantly less decline with age than will the conventional tests.

4. The tests which show least cultural bias will show substantial correlation with school achievement. Since these

culture-reduced tests are expected to show less educational loading it is predicted that they will correlate somewhat less with achievement than do conventional tests. In order to test this hypothesis "substantial correlation" is defined as a correlation of .40 or higher, provided such a correlation is significant at the .05 level.

CHAPTER IV

EXPERIMENTAL DESIGN AND PROCEDURE

Data Available From Previous and Current Studies

This investigation was designed to utilize data which were available from previous and current studies at the University of Alberta. The major data were collected in 1961 by Dr. R.S. McArthur for a factor analytic study. On the basis of previous research, he selected a battery of tests which included several conventional tests as well as several promising culture-reduced tests in order to provide for the identification of group factors and to permit comparisons between the conventional and the culture-reduced varieties.

At Faust, the battery of tests was administered in May 1961 by the home room teachers, under the supervision of Dr. MacArthur. All public school children in attendance were given the tests. At Fort Simpson the battery was similarly administered in October 1961. In addition to test scores, Dr. MacArthur also collected a great deal of data on other relevant variables.

Since each of the tests of the battery was standardized by the publisher on a different norming group, differences in performance on given tests may be attributed to differences in the abilitiesof the norming groups.

To exclude this interpretation and to permit the interpretation of such differences as evidence of cultural bias it was necessary to re-standardize each of the tests of the battery at a given grade level on a common norming group. Data collected for two other investigations were used to accomplish this task.

Data collected by Dr. W.B. Elley for his Ph. D. dissertation permitted several tests of the batteries administered at the Grade II and III and the Grade VII and VIII levels to be normed on a sample of Edmonton school children. Similar data collected by Dr. C. Safran and Mr. M. MacLean permitted several tests of the batteries administered at the Grade I, Grade II and III, and Grade V and VI levels to be normed on a sample of Calgary school children.

The scope of this investigation was for some purposes limited by the nature and extent of the data available. In a few instances it was necessary to employ statistical techniques to make available data comparable.

The Samples Studied

From the data collected by Dr. MacArthur at Faust and Fort Simpson two samples were available for investigation.

The Faust sample consisted of all Metis children in Grades 1, 2, 3, 5, 6, 7, and 8 who were attending the Faust

School at the time of test administration. Faust is a small village with a population of about 950. It is located approximately 225 miles north-west of Edmonton. Most of the Metis children attending the Faust school, however, live in the rural areas surrounding the village. The socio-economic status of the Metis families in the sample is extremely low as indicated by a mean Blishen Index of 38 and a mean Home Index of 4.1. Family life among the group is quite unstable. Common-law marriages of limited duration are not unusual. Many of the children of the sample are bilingual. In addition to English, Cree and some French are spoken in the majority of homes.

The sample consisted of a total of 126 pupils and was subdivided for study as follows:

(a) Grade I Level, (mean age: 7 yrs.9 mos.) N = 32

(b) Grade II and III Level, (mean age: 9 yrs.4 mos.) N = 42

(c) Grade V and VI Level, (mean age: 12 yrs.11 mos.) N = 29

(d) Grade VII and VIII Level, (mean age: 14 yrs.5 mos.) N = 23

The Fort Simpson sample permitted a complete replication of the investigation carried out with the Faust sample. It consisted of all Metis and Indian children in Grades 1, 2, 3, 5, 6, 7, 8, and 9 who were attending the Fort Simpson School at the time of test administration. Fort Simpson is a settlement on the Mackenzie River in the North West Territories at approximately 121° west longitude and 61½° north latitude. It is situated about 300 miles south of the Arctic Circle. As in the case of the Faust sample, the

socio-economic status of the families in the Fort Simpson sample was extremely low. The mean Blishen Index was 36. The Home Index was not obtained. The majority of pupils were also bilingual.

A total of 155 pupils were subdivided to match the Faust groupings as follows:

(a) Grade I Level, (mean age: 7 yrs. 6 mos) N = 19

(b) Grade II and III Level, (mean age:10 yrs. 2 mos) N = 46

(c) Grade V and VI Level, (mean age:13 yrs.11 mos) N = 58

(d) Grade VII and VIII Level, (mean age:13 yrs.10 mos) N = 32

An occasional student for which complete data were not available had to be eliminated from the study. It was assumed that the elimination of these few students would not significantly alter the representativeness of the sample.

The few grade nine students at Fort Simpson were included in the Grade VII and VIII Level. Their inclusion is justified by the fact that the battery of tests were administered to the Fort Simpson sample much earlier in the school term than was the case in Faust.

It was necessary to eliminate Grade IV students from the study since appropriate norms were not available. Grade V and VI norms were based on T-scores for Calgary Grade VI pupils. As one would expect, a majority of Faust and Fort Simpson Grade IV scores fell well below the lower limit of these norms. Grade III norms were based on a

different battery of tests and therefore also were inappropriate for use with Grade IV.

The Tests Investigated.

Each of the tests in the experimental battery (with the exception of SCRIT) is well known and commonly used in psychology and education for the assessment of intellectual ability. Each has content and face validity in the sense that a logical inspection of the test items leads one to believe that the behavior required for successful performance may be reasonably regarded as intelligent.

A brief description of each test and a summary of the evidence regarding its reliability and concurrent validity follows. Except at the Grade VII and VIII level, little evidence is available regarding the g loading of the tests. Such evidence is expected to arise from a factor analytic study for which the data was collected and which is presently being carried out by Dr. R.S. MacArthur.

The preliminary findings of Dr. MacArthur's study suggest that each of the tests, whether conventional or culture-reduced, by and large, measure the same sort of general ability. The g loading or construct validity of each test will be provided by Dr. MacArthur's factor analysis of the tests. On the basis of his preliminary findings the construct validity of the tests will be assumed for the present. The emperical validity of the tests however, will

be considered. For purposes of this investigation a test will be judged a valid measure of intellectual ability if it shows a significant and substantial correlation with school achievement.

A. The Grade I Battery.

1. Lorge-Thorndike, Level I (1954) is suitable for Kindergarten and Grade I children. The test consists of three subtests: Oral Vocabulary, Cross-Out (the one that does not belong), and Pairing (the two that go together). It is entirely pictorial and diagramatic. Although it does not require reading, a good comprehension of oral language is necessary, especially for Part I. The alternate forms reliability coefficient for the total test is given in the Technical Manual (1957) as .81. An odd-even reliability coefficient of .92 is also reported. Correlations with other intelligence tests are reported of the order .4 to .7. Evidence on the g loading, predictive and construct validity are lacking. Since the test as a whole is rather dependent upon oral language, it cannot be classified a priori, as culture-reduced.

2. The Colored Progressive Matrices (1947) consists of three sets of twelve matrices. Each matrix or pattern of four colored figures contains one omission. The examinee is required to choose from a series of six diagrams set out

below the matrix the one which is necessary to complete the matrix. The test is largely perceptual. It was designed for use with children about five to eleven years of age. It is unspeeded, easily administered, and intrinsically interesting. It purports to measure present capacity for intellectual activity irrespective of acquired knowledge. Burke (1958) reports re-test reliability coefficients of the order .8 for nine-year olds. Martin and Wiechers (1954) obtained a correlation of .91 with the WISC intelligence quotient for a group of Grade III pupils. For a sample of 300 Edmonton Grade III boys, MacArthur (1960) found the Colored Progressive Matrices test to have a high loading on G and no loadings on group factors. The test therefore appears to be one of the most promising culture-reduced instruments available.

3. The Safran Culturally Reduced Intelligence Test, SCRIT, (1960) consists of 36 items of increasing difficulty. Each item is composed of a series of colored geometrical figures which form a pattern according to some principle. One of the figures in the series is omitted. The task of the examinee is to select from a set of response options the one which completes the given series. The test is somewhat similar to the Colored Progressive Matrices, except that the figures form a series rather than a matrix. In the SCRIT, color is also a significant factor in the

principle upon which the series is based, whereas in the Matrices color is irrelevant and is employed only to increase interest and enhance the attractiveness of the test. Evidence regarding reliability, validity, and g loading are as yet lacking.

4. The Detroit Beginning First-Grade Intelligence Test (1937) was designed to assist in the classification of children entering the first grade. The test consists of ten parts with traditional types of items - marking the biggest object in a series of similar objects, identifying and remembering several objects from a set of difficult objects, similarities, identification of objects by use, supplying missing parts, locating absurd features in objects, identification of a class of objects from a set, copying diagrams, and following directions.

Engel and Baker (1937) report a correlation of .76 with the Stanford-Binet. They also obtained a split-half reliability coefficient of .91 based on a sample of 116 first grade pupils.

B. The Grade II and III Battery.

1. Lorge-Thorndike, Level 2, (1954) is similar to the Lorge-Thorndike, Level 1, although somewhat more difficult. An alternate form reliability coefficients for the total test is reported in the Technical Manual (1957) as .76.

As an indication of validity, correlations of .56 with the Kuhlmann-Anderson and .60 with the Otis at the Grade III level are also reported.

2. The Colored Progressive Matrices (1947)

See page 37.

3. SCRIT (1960): see page 38.

4. California Short-Form Test of Mental Maturity, (CTMM) Primary, (1953) consists of a set of seven sub-tests designed to measure four factors developed from logical rather than factorial constructs.

1.	Sensing right and left)	
2.	Manipulation of areas)	Spatial Relations
3.	Similarities)	
4.	Inference)	Logical Reasoning
5.	Number Series)	
6.	Numerical quantity)	Numerical Reasoning
7.	Vocabulary	Verbal Concepts

Separate scores may also be obtained for language and non-language factors.

It was the intention of the test designers to measure most of the types of mental processes sampled by the Binet scale. The CTMM therefore has a fairly high verbal loading. Language factors according to the manual account for about 37% of the total variance of the test. Non-spurious correlations between test factors range from .2 to .6 indicating the probability of a fairly large g factor.

The following split-half reliability coefficients are reported in the test manual.

Total	.92
Language	.88
Non-Language	.90
Spatial	.87
Logical	.87
Numerical	.85
Verbal	.82

The test requires about an hour of administration time. It is fairly easy to give and special "scoreze" sheets make scoring quick and accurate.

C. The Grade V and VI Battery.

1. Lorge-Thorndike, Non-Verbal, Level 3 (1954) consists of three sub-tests: Figure Analogies, Figure Classification, and Number Series. An inspection of test items suggests very little dependence upon specific past learning. The test is classified as culture-reduced since its items, by and large, could be solved in a variety of languages or modes of expression. Facility with numbers, however, could improve an individual's score on "Number Series" and hence limits the possibilities of the test as a culture-reduced instrument. Although it is administered with a definite time limit, it is essentially a power test. An alternate forms reliability coefficient for the total test of .81 is reported in the General Manual. As Odd-even reliability coefficient of .94 is also reported. As an indication of validity, correlations of .69 with the CTMM, .62 with

*
*
*
*
*
*
*

Kuhlmann-Anderson, and .66 with Otis tests are also reported.

Odd-even reliabilities for the sub-tests are given as follows:

Figure Classification	.71
Figure Analogies	.87
Number Series	.84

2. The Standard Progressive Matrices (1956)

See page 43.

3. IPAT, Cattell Test of g, Scale 2, (1949) was originally designed as a culture-free test of the g factor. The test is largely perceptual in nature and consists of forty-six items grouped into four sub-tests: Series, Classifications, Matrices, and Topology. It is not dependent upon either reading or culturally loaded pictures. The test is easy to administer, economical, and can be given in a period of less than twenty minutes. It is appropriate for children 8 to 13 years of age and for average unselected adults.

Cattell (1958) reports a retest reliability of .85 and split-half coefficients ranging from .7 to .9 for different samples. Buros' Mental Measurements Year Book (1958) reports a correlation with the Binet of .56, with the Otis of .73, and with the A.C.E. of .59. In a factor analysis of eleven tests Elley (1961) obtained a g loading of .79 for the Cattell.

4. SCRIT. See page 38.

D. The Grade VII and VIII Battery.

1. Lorge-Thorndike, Non-Verbal, Level 4, (1954) is similar to the Level 3 edition, although somewhat more difficult. An alternate forms reliability coefficient for the total test is given in the Technical Manual (1957) as .78. An odd-even reliability coefficient of .93 is also given. For a sample of Grade VII pupils a correlation of .83 is reported with the CTMM and a correlation of .85 with the Otis. For a sample of 271 Edmonton Grade VII students Elley (1961) carried out a factor analysis of eleven tests and obtained a g loading of .74 and a numerical factor loading of .33 for the total test. In a similar twenty-one variable factor analysis he obtained a g loading of .58 for Fig. Class., .55 for No. Series, .74 for Fig. Analogies, and a numerical factor loading of .40 for No. series.

2. Ravens Standard Progressive Matrices (1956) is similar to the Colored Progressive Matrices and consists of sixty items grouped into five sets of twelve. Each set requiresa different principle for solution. The initial item of a set is quite easy and self-explanatory. Succeeding items get progressively more difficult permitting the test to teach a complex principle of organization and to assess the individual's effectiveness at learning. The

Standard Matrices consists of patterns of nine figures, one of which is omitted. The examinee must choose from six or eight response options the figure which completes the pattern.

Several investigators have reported g loadings of the order .8. With a sample of 271 Edmonton Grade VII students Elley (1961) obtained a g loading of .78 and no other factor loadings. Raven (1960) reports re-test reliabilities ranging from .83 to .93 depending upon age. He also reports a correlation of .86 with the Terman-Merrill scale.

3. IPAT, Cattell Test of g, Scale 2, (1949)
See page 42.

4. California Short-Form Test of Mental Maturity, (CTMM), Elementary, 1957 is similar to the Primary form in that it presents the same type of problems under the same sub-test headings but at a more difficult level. Elley (1961) reports that at this level the CTMM is somewhat more dependent upon reading skills whereas at the Primary level more importance is attached to listening for instructions.

Reliability coefficients for the various sub-tests as given in the manual (1957) are set out below:

Test	K.R.(Gr.5)	Split-Half (Gr.4-6)
Total	.88	.95
Language	.83	.95
Non*Language	.80	.91
Spatial Relationships	.82	.87
Logical reasoning	.79	.87
Numerical reasoning	.60	.90
Verbal concepts	.76	.93

Validity coefficients corrected for attenuation on both test and criterion are reported in the manual. For the Language section these coefficients range from .8 to .9. For the Non-Language section they range from .6 to .7.

In a factor analysis of twenty-one variables on a sample of 271 Edmonton Grade VII students, Elley (1961) obtained the following factor loadings:

(a) CTMM - Number: g loading .64; v:ed loading .20; numerical facility loading .33.

(b) CTMM - Verbal: g loading .46; v:ed loading .66.

(c) CTMM - Spatial: g loading .61.

(d) CTMM - Logical: g loading .66.

On a similar factor analysis of eleven tests the same investigator obtained a g loading of .58 on CTMM Language with a v:ed loading of .62, and a g loading of .62 on CTMM Non-Language with an n loading of .35.

Analysis of Results.

In order to test Hypothesis I, raw scores on each test for the Faust and Fort Simpson Grade I samples were converted to derived scores based on Calgary Grade I T-scores. Raw scores for the Grade II and III samples were similarly based on Calgary Grade III T-scores and those for the Grade V and VI samples were based on Calgary Grade VI T-scores. For the Grade VII and VIII samples raw scores were converted to derived scores based on Edmonton Grade VII T-scores. Since all derived scores at a given level were based on the same standardization group, it was possible to compare the performance of the samples on different tests of the experimental battery. Differences in performance on the several tests relative to that of the urban white standardization group were interpreted as indicative of the extent of cultural bias in the tests.

The mean, standard deviation, and correlation with each of the other variables were found for all tests. For these computations, the I.B.M. 1620 was used. From the resulting data a comparison between the means of the various tests was made. The procedure was to submit each difference between means to a t-test using the standard error of the difference between correlated means as described by Garrett (1958), page 226. Although it was hypothesized that certain tests, a priori categorized as culture-reduced, would show

less bias than other tests, similarly classified as conventional, it was considered desirable for practical use to identify all test which showed significantly less bias than others, regardless of classification. Hence, a two-tailed test was necessary.

Calgary T-score norms for the subtests of the CTMM at the Grade II and III level were not available. Edmonton T-scores were available for these tests, although not for other tests of the battery. Since both Edmonton and Calgary T-score scales were available for the CTMM Total test, it was possible to predict the mean Calgary sub-test T-scores from the mean Edmonton sub-test T-scores. The assumption made was that the regression of the CTMM sub-test Calgary T-scores on the CTMM Total Calgary T-scores was equal to the regression of the corresponding CTMM sub-test Edmonton T-scores on the CTMM Total Edmonton T-scores. The method was an adaptation of that described by Guliksen (1950), page 299. Instead of predicting each score individually, then finding the mean of the predicted scores, the Edmonton T-score mean was treated as a single score and the Calgary T-score mean predicted directly. It is suggested that the regression effect which occurs in making predictions from a regression equation was minimized by this alteration.

For the Grade I level, and the Grade V and VI level seven different means resulted in twenty-one comparisons. However, at the Grade II and III level and the Grade VII

and VIII level thirteen means resulted in seventy-eight comparisons. When such a large number of comparisons are made some doubt attaches to the repeated use of the t-test. About 5 percent of the differences of such a large group may be expected to show significance by chance alone, when the .05 level of significance is used. Thus the more stringent .01 level of significance was chosen.

At the Grade VII and VIII level where it was necessary to make a large number of comparisons and where derived scores were available for each pupil on each variable, a statistically more rigorous method was used. A two-way analysis of variance was carried out and the error mean square found. Following the analysis of variance, Duncan's new multiple range test, as described by Edwards (1950), page 136, was made. Although an equal number of comparisons were required for the Grade II and III data, predicted means but not individual scores were available for the CTMM subtests, hence the procedure described above was not possible. Moreover, since the data at this level included several estimated values, it probably did not justify a more rigorous treatment. At the two levels where only twenty-one comparisons were made and the .01 level of significance was used, the chances are only about one in five that a single Type I error has occurred in testing all differences.

Differences in the mean performance on all tests at each grade level were tabulated for both samples. All

differences between means which are significant at the .01 level have been indicated by an asterisk. Since inferior performance on one test as compared to another is assumed to result from cultural bias in the former test, those tests on which the Faust and Fort Simpson samples obtain significantly lower means correspondingly possess significantly greater cultural bias.

Hypothesis II was tested at the Grade II and III level and also at the Grade VII and VIII level. Each test administered at these levels was classified by logical inspection prior to analysis of data. Tests which appeared to consist mainly of items that could be solved in any language or mode of expression and which are probably as novel to one cultural group as to another were classified as culture-reduced. Tests which did not appear to satisfy the above description were classified as conventional. Scores for all tests in each classification were combined and the mean and standard deviation of the combined groups were found. The difference between the means of the two groups was then submitted to the traditional t-test. A one-tailed test was used since the direction of the difference was hypothesized. Performance which was significantly higher on one group of test than on the other was again interpreted as indicating significantly less cultural bias in the former group of tests.

Some difficulty was encountered in testing Hypothesis

III as only cross-sectional data were available. Furthermore, the T-score norms for the battery of tests at the Grade II and III level and the Grade VII and VIII level were based on different groups. Hence, changes from one level to the other could be attributed to differences in ability of the samples at each level or to differences in the norms employed. It was necessary to assume that the shape of the profile of test means for the present Grade II and III class could be substituted for that of the Grade VII and VIII class five grades earlier. Considering the similarity in profile shape for the two widely separated Faust and Fort Simpson samples, similarity in shape of profile for two groups of children raised in the same community does not at all appear unreasonable. No assumptions were made regarding the equivalence in ability of the classes at the two different levels.

From a comparison of profile shapes only (see Figure V), it was obvious that whatever increase or decline in performance occurred on the culture-reduced tests over the five-grade interval, the corresponding difference in change in the conventional tests was of the opposite direction to that hypothesized. Since the direction and extent of change could not be determined from the cross-sectional data based on two sets of norms, it was assumed that there would be no mean change in the culture-reduced group of means. A linear transformation was made on the Grade II and III

test means such that the mean of the culture-reduced group of means would be the same as that of the Grade VII and VIII culture-reduced group of means. This procedure made it possible to investigate the extent and direction of change in performance on the conventional tests over the five-grade interval when the mean performance on the culture-reduced tests was held constant.

To accomplish the task, the change in resulting T-score means for each test, from the Grade II and III level to the Grade VII and VIII level, was found. These changes were treated as scores. The mean and standard deviation of the changes were then calculated for both the conventional and culture-reduced groups. The difference between mean change of each group was tested for significance by the method described by Walker and Lev (1953), page 158.

Hypothesis IV was tested by finding the correlation between all variables at each level and grade placement on the California Achievement Battery for that level. All correlations were tested for significance by reference to the table provided by Garrett (1958), page 201.

CHAPTER V

EXTENT OF BIAS IN VARIOUS TESTS

As previously stated, the major objective of this investigation is to identify test instruments which, for the two samples studied, show a minimum of cultural bias. Probably the most relevant findings are thus reported in this chapter.

To facilitate comparisons both between tests and between samples, the results are reported one grade level at a time with the Fort Simpson results (treated as replication findings) immediately following the Faust results.

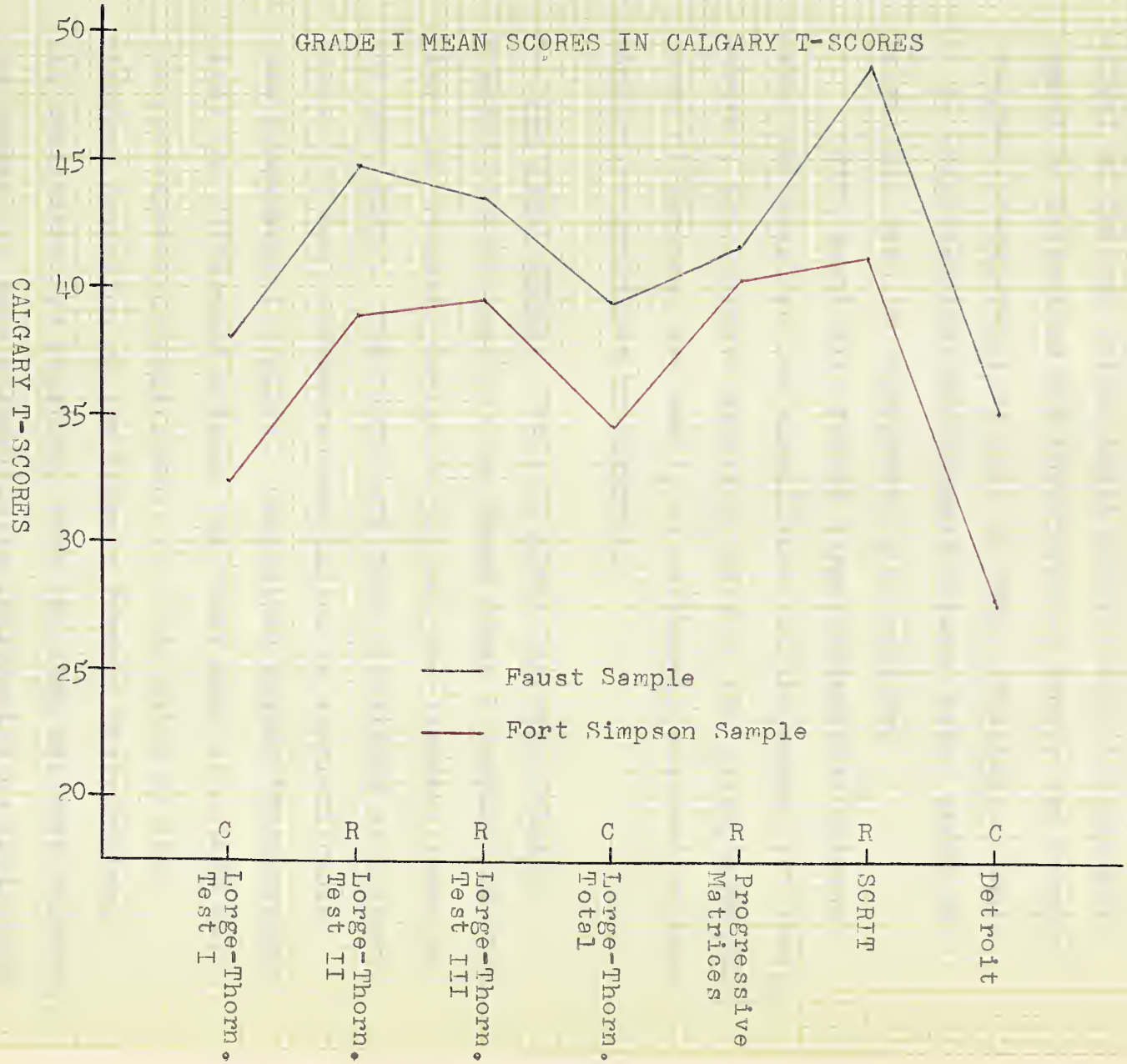
The Grade I Level.

Figure I shows the mean Calgary T-score of both the Faust and Fort Simpson Grade I samples on three conventional and four culture-reduced measures. Tests classified as conventional are indicated by a "C" whereas those classified as culture-reduced are indicated by an "R". The classificatory symbols are found just above the horizontal axis.

Since the deviation of the sample mean from the Calgary T-score mean of 50 is taken as an index of cultural bias in the test, those measures which yield means closest to 50 have, by definition, for that sample, least cultural b ias.

FIGURE I

GRADE I MEAN SCORES IN CALGARY T-SCORES



Calgary T-scores are based on raw scores for each variable except for the Detroit Beginning First Grade test. For this test Calgary T-scores are based on I.Q. The effect of the age factor which enters into this measure cannot be determined and therefore may render the T-score assigned incomparable to that of other variables. Thus, it is suggested that differences between means which involve this test be interpreted with caution.

With small N's, fairly large differences between means are required for significance at the chosen .01 level. However, the relative extent of bias in the tests, even when these differences are small, is strikingly similar for both samples as indicated by Figure I.

A. <u>The Faust Study</u>: Table I, gives the mean Calgary T-score of each test for the Faust Grade I sample. It also gives the standard deviation (S) and the standard error of the mean (SE_m). The tests have been classified as conventional (C) and culture-reduced (R) on an a priori basis (see Hypothesis II, p.29). The column headed "σ-deviation" gives the difference between the Faust mean of a test and the corresponding Calgary mean of 50 in units of the standard deviation of the Calgary T-score distribution. All deviations are negative, thus the sign has been omitted. The numbers in this column may be interpreted as indicating the relative extent of cultural bias in the tests.

TABLE I

FAUST GRADE I TEST BATTERY WITH MEASURES BASED ON CALGARY GRADE I T-SCORES N = 32

Test or Subtest	*	Mean	S	SE_m	σ-Dev.
Lorge-Thorndike I (Oral Vocabulary)	C	37.5	8.4	1.485	1.25
Lorge-Thorndike II (Cross-Out)	R	44.5	8.2	1.450	.55
Lorge-Thorndike III (Pairing)	R	43.4	9.8	1.733	.66
Lorge-Thorndike Total	C	39.4	7.7	1.361	1.06
Coloured Progressive Matrices	R	41.4	6.9	1.220	.86
S C R I T	R	48.5	5.9	1.043	.15
Detroit Beginning I.Q.	C	35.0	4.2	.743	1.50

* Classification: C - conventional. R - culture-reduced.

Table II presents a matrix of all possible differences between the means of the tests for the Faust Grade I sample.

TABLE II

DIFFERENCES BETWEEN MEANS: FAUST SAMPLE, GRADE I

Test or Subtest	L.Th. I	L.Th. Total	Prog. Mat.	L.Th. III	L.Th. II	SCRIT
Detroit Beg.	2.5	4.4*	6.3*	8.4*	9.5*	13.5*
L.Th. I		1.9	3.8	5.9*	7.7*	11.0*
L.Th. Total			1.9	4.0*	5.1*	9.1*
Prog. Mat.				2.1	3.2	7.2*
L.Th. III					1.1	5.1*
L.Th. II						4.0*

* Significant at the .01 level.

The results summarized in Tables I and II support the first two hypothesis for the battery of tests administered to the Faust Grade I sample. Some of the tests of the battery have significantly less cultural bias than do others. Furthermore, those which show a minimum of cultural bias are, without exception, tests which prior to analysis have been classified as consisting largely of items that can be solved in any language or mode of expression and which are probably as novel to one culture as to another.

B. The Fort Simpson Replication: Table III gives the mean Calgary T-score for each variable in the Fort Simpson Grade I sample. It also gives the standard deviation (S), the standard error of the mean (SE_m), the classification of the test (*) and the "δ-deviation" as described on page 54 .

TABLE III

FORT SIMPSON GRADE I BATTERY WITH MEASURES BASED ON CALGARY GRADE I T-SCORES N = 19

Test or Subtest	*	Mean	S	SE_m	σ-Dev.
Lorge-Thorndike I (Oral Vocabulary)	C	32.4	7.0	1.606	1.76
Lorge-Thorndike II (Cross-Out)	R	38.9	10.3	2.363	1.11
Lorge-Thorndike III (Pairing)	R	39.5	9.0	2.065	1.05
Lorge-Thorndike Total	C	34.5	6.0	1.376	1.55
Coloured Progressive Matrices	R	40.5	20.9	4.794	.95
S C R I T	R	41.4	9.8	2.248	.86
Detroit Beginning I.Q.	C	27.4	7.6	1.743	2.26

* Classification: C - conventional; R - culture-reduced

Table IV presents a matrix of all possible differences between the means of the tests for the Fort Simpson Grade I sample.

TABLE IV

DIFFERENCES BETWEEN MEANS: FORT SIMPSON SAMPLE, GRADE I

Test or Subtest	L.Th. I	L.Th. Total	L.Th. II	L.Th. III	Prog. Mat.	SCRIT
Detroit Beg.	5.0	7.1*	11.5*	12.1*	13.1*	14.0*
L.Th. I		2.1	6.5*	7.1*	8.1	9.0*
L.Th. Total			4.4	5.0*	6.0	6.9
L.Th. II				0.6	1.6	2.5
L.Th. III					1.0	1.9
Prog. Mat.						0.9

* Significant at the .01 level.

The results summarized in Tables III and IV support the first two hypothesis for the battery of tests administered to the Fort Simpson Grade I sample.

<u>The Grade II and III Level.</u>

Figure II shows the mean Calgary T-score of both the Faust and the Fort Simpson Grade II and III samples for thirteen different variables. To facilitate comparisons, these measures have been grouped according to the a priori

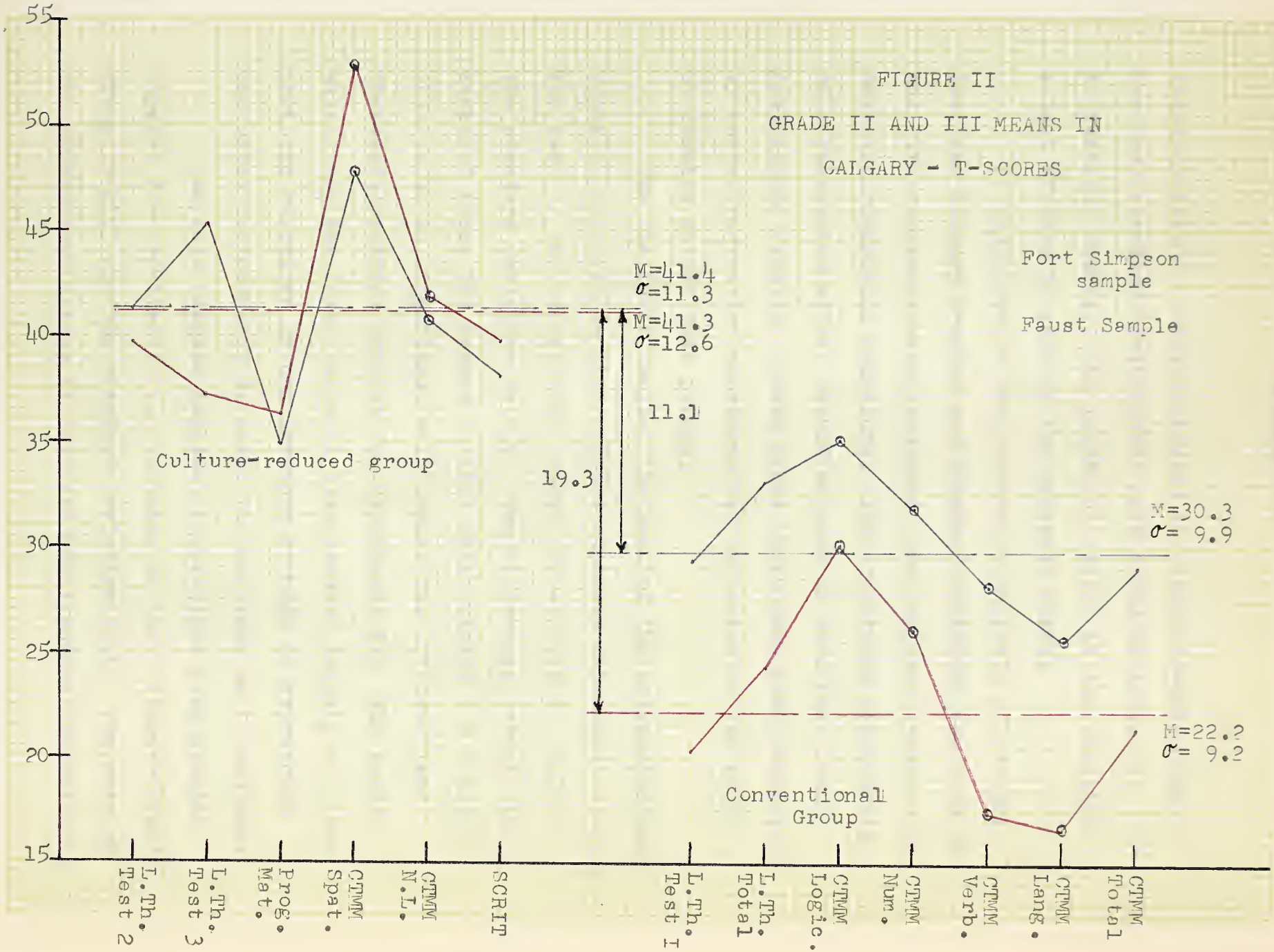


FIGURE II

GRADE II AND III MEANS IN

CALGARY - T-SCORES

classification: culture-reduced and conventional. The mean and standard deviation of each group of tests is indicated. However, the important factor is the relative extent of bias in each of the separate tests.

As explained in the section on analysis of results, the mean Calgary T-score and standard deviation for each of the CTMM sub-tests were predicted from Edmonton T-scores by means of regression equations. The consequent possibility of a regression effect should be kept in mind when interpreting the results. These means which have been obtained by use of regression equations are indicated on the graph by circles rather than points.

For the Faust sample, the mean of the culture-reduced group of measures is 41.4 and the standard deviation is 11.3. The mean of the conventional group of measures is 30.3 and the standard deviation is 9.9. The difference between the means of these two groups is 11.1 which yields a t = 12.1. This value is significant well beyond the .01 level and therefore provides support for Hypothesis II. The tests which show the least cultural bias consist largely of items that can be solved in any language or mode of expression and which are probably as novel to one group as to another.

The Fort Simpson replication provides even greater support for Hypothesis II. The mean of the culture-reduced group is 41.3 and the standard deviation 12.6. The mean of the conventional group is 22.2 and the standard deviation is

9.2. The difference between the means of the two groups is 19.3 which yields a t = 21.1, again significant well beyond the .01 level.

It is interesting to note that whereas the culture-reduced tests yield means which are negligibly different for the two samples, the conventional tests yield means which are significantly different (t = 10.5). Not only do the verbal or conventional group of tests show significantly greater bias for both samples than do the culture-reduced group, but they also show significantly greater bias toward the more remote Fort Simpson sample than toward the Faust sample.

A. <u>The Faust Study</u>: Table V gives the mean Calgary T-score of each variable for the Faust Grade II and III sample. It also gives the standard deviation (S), the standard error of the mean (SE_m), and the "δ-deviation" as explained on page 54 . All statistics based on predicted measures are underlined.

TABLE V

FAUST GRADE II AND III TEST BATTERY WITH MEASURES BASED ON CALGARY GRADE III T-SCORES N = 42

Test or Subtest	*	Mean	S	SE_m	σ-Dev.
Lorge-Thorndike I (Oral Vocabulary)	C	29.4	7.2	1.111	2.06
Lorge-Thorndike II (Cross Out)	R	41.3	8.3	1.281	.87
Lorge-Thorndike III (Pairing)	R	45.2	10.1	1.558	.48
Lorge-Thorndike Total	C	33.0	9.6	1.481	1.70
Coloured Progressive Matrices	R	35.0	11.1	1.713	1.50
S C R I T	R	38.4	8.7	1.342	1.16
CTMM - Total	C	29.0	11.5	1.774	2.10
- Spatial	R	47.6	11.5	1.774	.24
- Logical	C	35.0	8.5	1.312	1.50
- Numerical	C	31.9	11.0	1.697	1.81
- Verbal	C	28.0	8.7	1.342	2.20
- Language	C	25.5	8.5	1.312	2.45
- Non-Language	R	40.8	12.8	1.975	.92

* Classification: C - conventional; R - Culture-reduced.

Table VI presents a matrix of all possible differences between the means of the tests for the Faust Grade II and III sample. Differences which are significant at the .01 level are indicated by an asterisk. Differences involving predicted measures are underlined.

A total of 78 differences have been tested and the results reported in Table VI. There is some question as to the advisability of testing so many differences. For this reason the rather stringent .01 level of significance was chosen. It is unlikely therefore that more than one difference has yielded a significant t as a result of chance alone.

Although it was possible to predict mean Calgary T-scores for the CTMM subtests, individual scores were not available. Hence it was impossible to employ the more rigorous statistical techniques which follow an analysis of variance such as the Duncan's Multiple Range Test.

The results summarized in Table VI, nevertheless, adequately support Hypothesis I. Some of the tests of the battery have, for the Faust Grade II and III sample, significantly less cultural bias than do others.

TABLE VI

Differences Between Means, Faust Grade 2 and 3 Battery in Calgary T-scores.

	CTMM Verb.	CTMM Total	L.Th. I	CTMM Num.	L.Th. Total	CTMM Log.	Prog. Mat.	SCRIT	CTMM Non-Lang.	L.Th. II	L.Th. III	CTMM Spat.
CTMM - Language	2.5	3.5*	3.9*	6.4*	7.5*	9.5*	9.5*	12.9*	15.3*	15.8*	19.7*	22.1*
CTMM - Verbal		1.0	1.4	3.9	5.0*	7.0*	7.0*	10.4*	12.8*	13.3*	17.2*	19.6*
CTMM - Total			.4	2.9	4.0	6.0*	6.0*	9.4*	11.8*	12.3*	16.2*	18.6*
Lorge-Thorndike Test I				2.5	3.6*	5.6*	5.6*	9.0*	11.4*	11.9*	15.8*	18.2*
CTMM Numerical					1.1	3.1	3.1	6.5*	8.9*	9.4*	13.3*	15.7*
Lorge-Thorndike Total						2.0	2.0	5.4*	7.8*	8.3*	12.2*	14.6*
CTMM - Logical							0.0	3.4	5.0*	6.3*	10.2*	12.6*
Progressive Matrices								3.4*	5.8*	6.3*	10.2*	12.6*
SCRIT									2.4	2.9*	6.8*	9.2*
CTMM Non-Language										.5	4.4	6.8*
Lorge-Thorndike Test II											3.9	6.3*
Lorge-Thorndike Test III												2.4

* Differences significant at the .01 level.

B. The Fort Simpson Replication: Table VII gives the mean Calgary T-score of each variable for the Fort Simpson Grade II and III sample. It also gives the standard deviation (S), the standard error of the mean (SE_m), and the "σ-deviation". All statistics based on predicted measures are underlined.

TABLE VII

FORT SIMPSON GRADE II AND III TEST BATTERY WITH MEASURES BASED ON GRADE III CALGARY T-SCORES N = 46

Test or Subtest	*	Mean	S	SE_m	σ-Dev.
Lorge-Thorndike I (Oral Vocabulary)	C	20.3	7.4	1.092	2.97
Lorge-Thorndike II (Cross Out)	R	39.8	9.8	1.446	1.02
Lorge-Thorndike III (Pairing)	R	37.2	14.1	2.080	1.28
Lorge-Thorndike Total	C	24.2	10.3	1.519	2.58
Coloured Progressive Matrices	R	36.3	12.2	1.800	1.37
S C R I T	R	39.8	11.1	1.637	1.02
CTMM - Total	C	21.3	8.0	1.180	2.87
- Spatial	R	52.9	10.2	1.505	-.29
- Logical	C	30.0	6.2	.915	2.00
- Numerical	C	26.0	10.3	1.519	2.40
- Verbal	C	17.4	7.4	1.092	3.26
- Language	C	16.5	5.6	.826	3.35
- Non-Language	R	41.8	10.2	1.505	.82

* Classification: C - Conventional: R - Culture-reduced

Table VIII presents a matrix of all possible differences between means of tests for the Fort Simpson Grade II and III sample. Differences which are significant at the .01 level are indicated by an asterisk. Differences which involve predicted measures are underlined. The results summarized in Table VIII provide additional support for Hypothesis I.

TABLE VIII

Differences Between Means, Fort Simpson Grade 2 and 3 Battery in Calgary T-scores

	CTMM Verb.	L.Th. I	CTMM Total	L.Th. Total	CTMM Num.	CTMM Log.	Prog. Mat.	L.Th. III	L.Th. II	SCRIT	CTMM Non-Lang.	CTMM Spat.
CTMM - Language	.9	3.8*	4.8*	7.7*	9.5*	13.5*	19.8*	20.7*	23.3*	23.3*	25.3*	36.4*
CTMM - Verbal		2.9	3.9*	6.8*	8.6*	12.6*	18.9*	19.8*	22.4*	22.4*	24.4*	35.5*
Lorge-Thorndike Test I			1.0	3.9	5.7*	9.7*	16.0*	16.9*	19.5*	19.5*	21.5*	32.6*
CTMM - Total				2.9	4.7*	8.7*	15.0*	15.9*	18.5*	18.5*	20.5*	31.6*
Lorge-Thorndike Total					1.8	5.8*	12.1*	13.0*	15.6*	15.6*	17.6*	28.7*
CTMM Numerical						4.0*	10.3*	11.2*	13.8*	13.8*	15.8*	26.9*
CTMM - Logical							6.3*	7.2*	9.8*	9.8*	11.8*	22.9*
Progressive Matrices								.9	3.5	3.5	5.5*	16.6*
Lorge-Thorndike Test III									2.6	2.6	4.6	15.7*
Lorge-Thorndike Test II										.0	2.0	13.1*
SCRIT											2.0	13.1*
CTMM Non-Language												11.1*

* Differences significant at the .01 level.

The Grade V and VI Level.

At this level the data were especially limited. Lack of data made it impossible to obtain comparable T-scores for the conventional tests. Therefore Hypothesis II could not be tested. Data did permit scaling of seven culture-reduced variables however. Figure III shows the mean Calgary T-score of both the Faust and Fort Simpson samples for these variables.

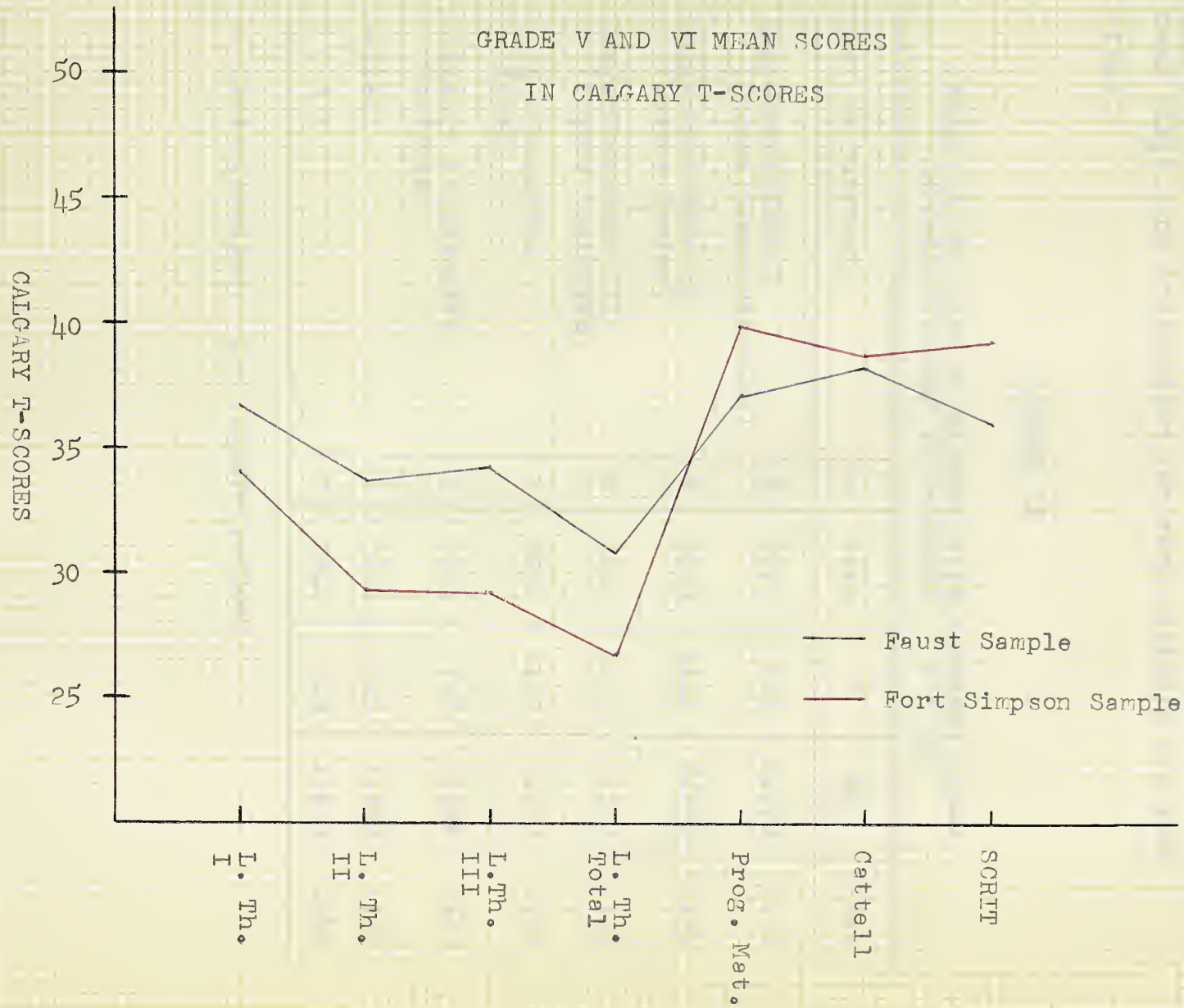
The analyses of data for both samples supports Hypothesis I. Even among tests classified a priori as culture-reduced some are, for the samples studied, less culturally biased than are others. Such evidence may have considerable practical value in selecting from among non-verbal or culture-reduced tests at this level.

It is interesting to note the similarity in the relative extent of cultural bias of the tests for the two samples as illustrated by Figure III. It is also noteworthy that whereas all of the Lorge-Thorndike Level 3 tests appear to show somewhat greater bias against the Fort Simpson sample than against the Faust sample, the three other measures appear to show less bias.

FIGURE III

GRADE V AND VI MEAN SCORES

IN CALGARY T-SCORES



A. The Faust Study: Table IX gives the mean Calgary T-score of seven variables for the Faust Grade V and VI sample. The standard deviation (S), standard error of the mean (SE_m), and σ-deviation for each variable are also given.

TABLE IX

FAUST GRADE V AND VI BATTERY WITH MEASURES BASED ON CALGARY GRADE VI T-SCORES N = 29

Test or Sub-test	*	Mean	S	SE_m	σ-Dev.
Lorge-Thorndike I (Figure Classification)	R	36.5	8.5	1.578	1.35
Lorge-Thorndike II (Number Series)	R	33.6	11.1	2.061	1.64
Lorge-Thorndike III (Figure Analogies)	R	34.1	15.7	2.915	1.59
Lorge-Thorndike Total	R	30.8	11.0	2.042	1.92
Standard Progressive Matrices	R	37.1	9.7	1.801	1.29
Cattell	R	38.4	10.5	1.950	1.16
S C R I T	R	36.0	8.9	1.653	1.40

* Classification: R - culture-reduced.

Table X gives the matrix of all differences between the means of the tests for the Faust Grade V and VI sample.

TABLE X

DIFFERENCES BETWEEN MEANS: FAUST SAMPLE, GRADE V AND VI

Test or Sub-test	L.Th. III	L.Th. II	L.Th. I	Cattell	SCRIT	Prog. Mat.
L.Th. Total	2.8	3.3	5.2*	5.7*	6.3*	7.6*
L.Th. II		.5	2.4	2.9	3.5	4.8
L.Th. III			1.9	2.4	3.0	4.3
S C R I T				.5	1.1	2.4
L.Th. I					.6	1.9
Prog. Mat.						1.3

* Significant at the .01 level.

B. The Fort Simpson Replication: Table XI gives the mean Calgary T-score of the seven variables for the Fort Simpson Grade V and VI sample. The standard deviation (S), standard error of the mean (SE_m), and σ-deviation for each variable is also given.

TABLE XI

FORT SIMPSON GRADE V AND VI BATTERY WITH MEASURES BASED ON CALGARY GRADE VI T-SCORES N = 58

Test or Sub-test	*	Mean	S	SE_m	σ-Dev.
Lorge-Thorndike I (Fig. Classification)	R	34.2	10.6	1.392	1.58
Lorge-Thorndike II (Number Series)	R	29.4	14.8	1.943	2.06
Lorge-Thorndike III (Fig. Analogies)	R	29.2	15.7	2.061	2.08
Lorge-Thorndike III Total	R	26.8	13.3	1.746	2.32
Standard Progressive Matrices	R	39.9	10.6	1.392	1.01
Cattell	R	38.5	11.5	1.510	1.15
S C R I T	R	39.2	9.0	1.182	1.08

* Classification: R - culture-reduced

Table XII gives the matrix of all differences between the means of the tests for the Fort Simpson Grade V and VI sample.

TABLE XII

DIFFERENCES BETWEEN MEANS: FORT SIMPSON SAMPLE, GRADE V AND VI

Test or Sub-test	L.Th. III	L.Th. II	L.Th. I	Cattell	SCRIT	Prog. Mat.
L.Th. Total	2.4	2.6	7.4*	11.7*	12.4*	13.1*
L.Th. III		.2	5.0*	9.3*	10.0*	10.7*
L.Th. II			4.8*	9.1*	9.8*	10.5*
L.Th. I				4.3*	5.0*	5.7*
Cattell					.7	1.4
S C R I T						.7

* Significant at the .01 level.

The Grade VII and VIII Battery.

Data for the Grade VII and VIII samples have been scaled in terms of T-scores based on a representative sample of 271 Edmonton Grade VII pupils. To illustrate the relative extent of cultural bias in the various tests, the mean Edmonton T-scores of both the Faust and the Fort Simpson samples for thirteen variables are plotted in Figure IV. In order to test Hypothesis II, each of these variables has been classified a priori as culture-reduced or conventional.

The mean and standard deviation of each of the two classes of tests taken as a group are also indicated in Figure IV.

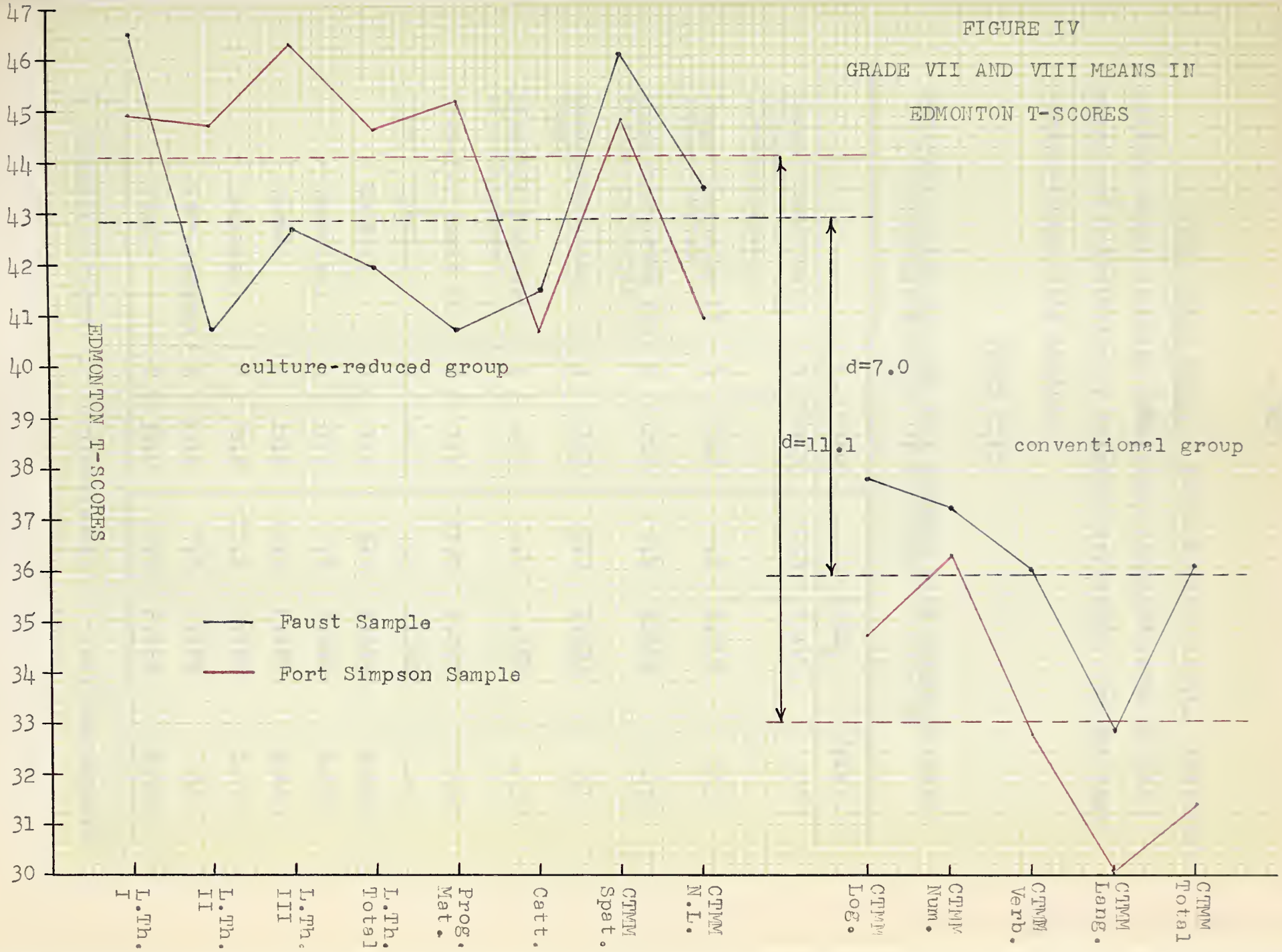
For the Faust sample the mean of the culture-reduced group of tests is 42.9 and the standard deviation is 8.9. The mean of the conventional group of tests is 35.9 and the standard deviation is 11.2. The difference between the means for the two groups of tests is 7.0 which is significant well beyond the .01 level ($t = 5.7$). The Faust data at this level therefore support Hypothesis II.

Even greater support for Hypothesis II is provided by the Fort Simpson replication. For this sample the mean of the culture-reduced group of tests is 44.1, and the standard deviation is 9.7. The mean of the conventional group is 33.0 and the standard deviation is 11.1 which again is significant well beyond the .01 level ($t = 10.4$).

FIGURE IV

GRADE VII AND VIII MEANS IN

EDMONTON T-SCORES



A. The Faust Study: Table XIII gives the classification, mean, standard deviation, standard error of the mean, and σ-deviation of thirteen variables, for the Faust Grade VII and VIII sample.

TABLE XIII

THE FAUST GRADE VII AND VIII BATTERY WITH MEASURES BASED ON EDMONTON GRADE VII T-SCORES N = 23

	*	Mean	S	SE_m	σ-Dev.
Lorge-Thorndike I Figure Classification	R	46.5	10.3	2.145	.35
Lorge-Thorndike II Number Series	R	40.7	6.8	1.416	.93
Lorge-Thorndike III Figure Analogies	R	42.7	9.9	2.062	.73
Lorge-Thorndike Total	R	41.9	8.8	1.833	.81
Standard Progressive Matrices	R	40.7	6.0	1.250	.93
Cattell, Scale II	R	41.4	11.0	2.291	.86
CTMM - Spatial	R	46.0	7.5	1.562	.40
- Logical	C	37.7	8.7	1.812	1.23
- Numerical	C	37.1	9.9	2.062	1.29
- Verbal	C	35.9	12.9	2.687	1.41
- Language	C	32.8	12.6	2.625	1.72
- Non-Language	R	43.4	7.8	1.625	.66
- Total	C	36.0	10.7	2.229	1.40

* Classification: C - conventional; R - culture-reduced

Table XIV presents a matrix of all differences between the means of the thirteen variables in the Grade VII and VIII battery for the Faust sample. Differences which are significant at the .01 level using a two-way analysis of variance followed by Duncan's new multiple range test are indicated by an asterisk. The column headed "Shortest Significant Range" gives the difference required for significance at the .01 level for each range.[1] At the chosen .01 level of significance the 13-mean protection level is .89. This figure gives the probability of no erroneous significant differences between the thirteen means. In other words, the probability that a Type I error has not been made in a single comparison is greater than .99, but the probability that a Type I error has not been made in the total matrix of 78 differences is approximately .89.

Tables XIII and XIV summarize evidence in support of Hypothesis I. Some of the tests of the experimental battery administered to the Grade VII and VIII Faust sample show significantly less cultural bias than do others. These tests have been identified.

---

[1]For a description of the Duncan's new multiple range test see Edwards, A.L. Experimental Design in Psychological Research. New York: Holt, Rinehart and Winston, 1960. p.136.

TABLE XIV

Differences Between Means of Tests; Faust Grade 7 and 8 Battery in Edmonton T-scores

Test or Subtest	CTMM Verb.	CTMM Total	CTMM Num.	CTMM Log.	Prog. Mat.	L.Th. II	Cattell	L.Th. Total	L.Th. III	CTMM Non-Lang.	CTMM Spat.	L.Th. I	Shortest Significant Range
CTMM - Language	3.1	3.2	4.3	4.9	7.9*	7.9*	8.6*	9.1*	9.9*	10.6*	13.2*	13.7*	$R_2 = 5.1$
CTMM - Verbal		.1	1.2	1.8	4.8	4.8	5.5	6.0*	6.8*	7.5*	10.1*	10.6*	$R_3 = 5.3$
CTMM - Total			1.1	1.7	4.7	4.7	5.4	5.9*	6.7*	7.4*	10.0*	10.5*	$R_4 = 5.4$
CTMM Numerical				.6	3.6	3.6	4.3	4.8	5.6	6.3*	8.9*	9.4*	$R_5 = 5.6$
CTMM - Logical					3.0	3.0	3.7	4.2	5.0	5.7*	8.3*	8.8*	$R_6 = 5.6$
Progressive Matrices						.0	.7	1.2	2.0	2.7	5.3	5.8*	$R_7 = 5.7$
Lorge-Thorndike Test II							.7	1.2	2.0	2.7	5.3	5.8*	$R_8 = 5.8$
Cattell								.5	1.3	2.0	4.6	5.1	$R_9 = 5.8$
Lorge-Thorndike Total									.8	1.5	4.1	4.6	$R_{10} = 5.9$
Lorge-Thorndike Test III										.7	3.3	3.8	$R_{11} = 5.9$
CTMM Non-Language											2.6	3.1	$R_{12} = 6.0$
CTMM - Spatial												.5	$R_{13} = 6.0$

Asterisks indicate differences which are significant at the .01 level, using twoway analysis of variance and Duncan's new multiple range test.

B. The Fort Simpson Replication: The classification, mean, standard deviation, standard error of the mean, and σ-deviation of thirteen variables for the Fort Simpson Grade VII and VIII sample are given in Table XV.

TABLE XV

THE FORT SIMPSON GRADE VII AND VIII BATTERY WITH MEASURES BASED ON EDMONTON GRADE VII T-SCORES N = 32

	*	Mean	S	SE_m	σ-Dev.
Lorge-Thorndike I Figure Classification	R	45.0	12.1	2.139	.50
Lorge-Thorndike II Number Series	R	44.8	8.7	1.538	.52
Lorge-Thorndike III Figure Analogies	R	46.4	8.5	1.503	.36
Lorge-Thorndike Total	R	44.7	9.2	1.627	.53
Standard Progressive Matrices	R	45.2	7.3	1.291	.48
Cattell, Scale II	R	40.8	10.9	1.927	.92
CTMM - Spatial	R	44.9	10.1	1.786	.51
- Logical	C	34.8	9.1	1.609	1.52
- Numerical	C	36.3	8.9	1.574	1.37
- Verbal	C	32.8	12.0	2.122	1.72
- Language	C	29.6	13.1	2.316	2.04
- Non-Language	R	41.0	8.4	1.485	.90
- Total	C	31.4	10.3	1.821	1.86

* Classification: C - conventional: R - culture-reduced.

Analogous to Table XIV of the Faust study, a matrix of all differences between the means of the thirteen variables for the Fort Simpson sample is presented in Table XVI. Differences which are significant at the .01 level using a two-way analysis of variance followed by Duncan's new multiple range test are indicated by an asterisk. The 13-mean protection level is again .89.

The findings summarized in Tables XV and XVI show that the Fort Simpson replication provides additional support for Hypothesis I. Several of the tests administered to the Fort Simpson Grade VII and VIII sample show significantly less cultural bias than do others. These tests also have been identified.

TABLE XVI

Differences Between Means of Tests; Fort Simpson Grade VII and VIII Battery in Edmonton T-scores

Test or Subtest	CTMM Total	CTMM Verb.	CTMM Log.	CTMM Num.	Cattell	CTMM Non-Lang.	L.Th. Total	L.Th. II	CTMM Spat.	L.Th. I	Prog. Mat.	L.Th. III	Shortest Significant Range
CTMM - Language	1.8	3.2	5.2	6.7*	11.2*	11.4*	15.1*	15.2*	15.3*	15.4*	15.6*	16.8*	R_2=5.2
CTMM - Total		1.4	3.4	4.9	9.4*	9.6*	13.3*	13.4*	13.5*	13.6*	13.8*	15.0*	R_3=5.4
CTMM - Verbal			2.0	3.5	8.0*	8.2*	11.9*	12.0*	12.1*	12.2*	12.4*	13.6*	R_4=5.5
CTMM - Logical				1.5	6.0*	6.2*	9.9*	10.0*	10.1*	10.2*	10.4*	11.6*	R_5=5.7
CTMM Numerical					4.5	4.7	8.4*	8.5*	8.6*	8.7*	8.9*	10.1*	R_6=5.7
Cattell						.2	3.9	4.0	4.1	4.2	4.4	5.6	R_7=5.8
CTMM Non-Language							3.7	3.8	3.9	4.0	4.2	5.4	R_8=5.9
Lorge-Thorndike Total								.1	.2	.3	.5	1.7	R_9=5.9
Lorge-Thorndike Test II									.1	.2	.4	1.6	R_{10}=6.0
CTMM - Spatial										.1	.3	1.5	R_{11}=6.0
Lorge-Thorndike Test I											.2	1.4	R_{12}=6.1
Progressive Matrices												1.2	R_{13}=6.1

Asterisks indicate differences which are significant at the .01 level using two-way analysis of variance and Duncan's new multiple range test.

CHAPTER VI

CHANGES IN EXTENT OF CULTURAL BIAS OVER AN INTERVAL OF FIVE GRADES

Hypothesis III states that the culture-reduced tests common to the experimental batteries for the Grade II and III level and the Grade VII and VIII level, will for the samples studied, show less decline with age than will the conventional tests.

If a stimulating cultural environment facilitates intellectual development whereas an impoverished cultural environment hinders intellectual development, it follows that the difference in intelligence between members of these two cultures will become increasingly greater with continued exposure to their respective environments. It was assumed that white urban children from Edmonton and Calgary are exposed to a more intellectually stimulating environment than are the Faust and Fort Simpson children. Thus, it was expected that, with increasing age of subjects, the Faust and Fort Simpson test scores would decline relative to Edmonton and Calgary scores. It has also been assumed that the culture-reduced test are less affected by environmental differences than are the conventional tests. Hypothesis III therefore arises from the postulate that tests dependent upon verbal symbolism and past specific learning experiences will increasingly discriminate against

culturally handicapped groups.

A rigorous test of Hypothesis III would require longitudinal data which are not available. The cross-sectional data at hand however, are worthy of consideration.

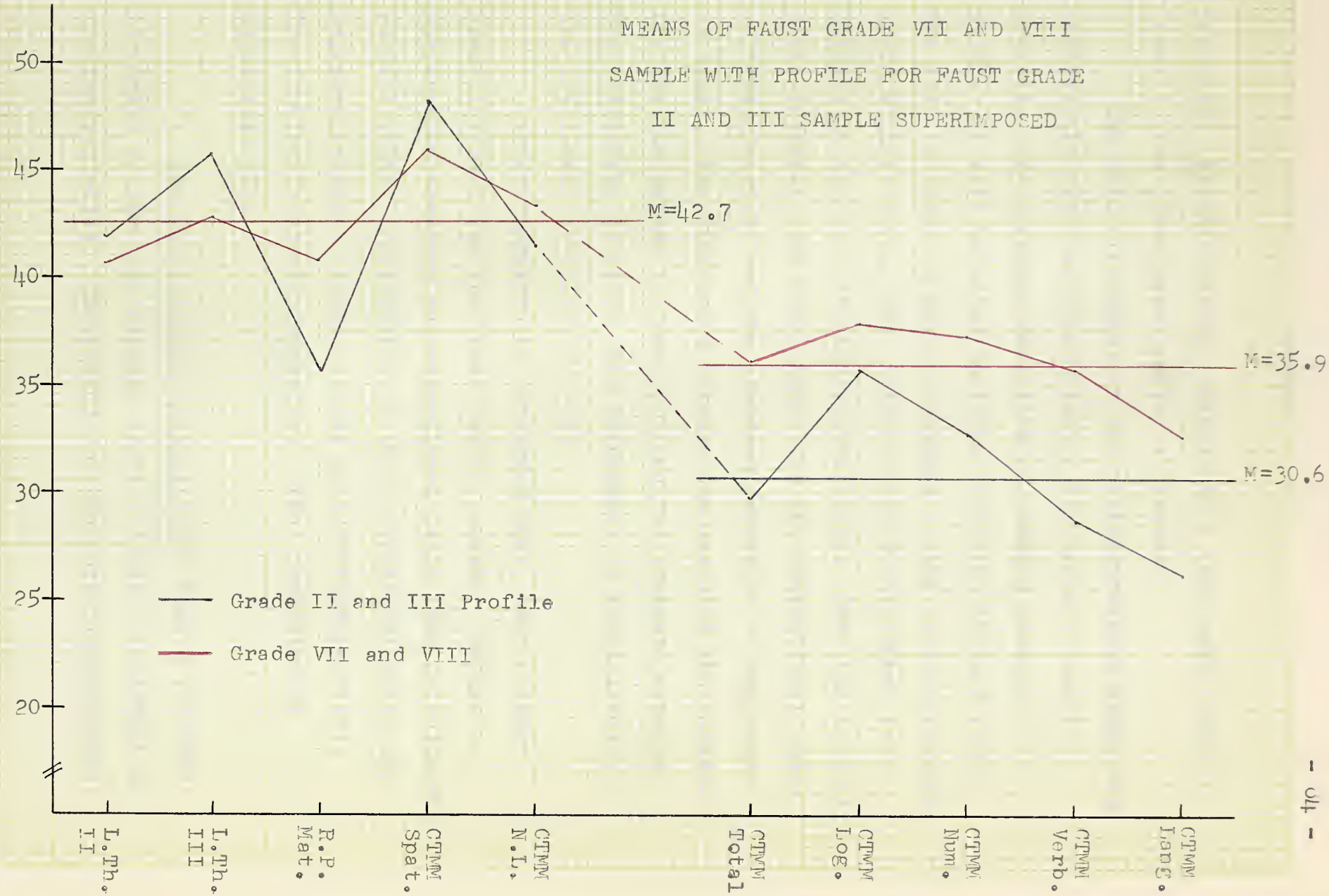
In order to investigate the extent and direction of change in performance on the conventional tests as compared with the culture-reduced tests over the five-grade interval it was necessary to make several assumptions. The assumptions made and the technique of analysis employed were described previously in the section entitled "Analysis of Results".

A. The Faust Sample: Figure V shows the means of the Faust Grade VII and VIII sample on five conventional and five culture-reduced variables. The profile of the Grade II and III sample is superimposed upon that of the Grade VII and VIII, such that the means of the culture-reduced group for both levels are equal. The mean of the conventional group of tests at each level is also shown.

It will be noted that when the mean of the culture-reduced group is equivalent at both levels, the mean of the conventional group increases over the five-grade interval from 30.6 to 35.9. The difference between the mean change in culture-reduced measures and the mean change in conventional measures over the interval was tested for significance using the method described by Walker and Lev (1953),

FIGURE V

MEANS OF FAUST GRADE VII AND VIII SAMPLE WITH PROFILE FOR FAUST GRADE II AND III SAMPLE SUPERIMPOSED



page 158. By this method, changes in a test mean from one level to the other were treated as scores.

The mean of changes in the culture-reduced group was arbitrarily made zero when placing the Grade II and III profile. The standard deviation of changes however, is 3.385. It will be noted that the condition imposed upon the data is that the mean of changes of the culture-reduced group will be zero. This does not imply that there are no changes in means of the individual tests. Over the investigated interval, the mean change for the conventional group of tests is +5.29 and the standard deviation of the change is 2.135. The difference between the means of the changes for the two groups is thus 5.29 and the standard error of the difference is 1.79. The difference is significant at the .05 level ($t = 2.96$, $df = 8$).

It may therefore be concluded that, over a five-grade interval for the Faust sample, greater changes in performance occurred on conventional tests than on culture-reduced tests. Performance on the conventional tests was improved relative to that of the culture-reduced tests. The finding is exactly opposite to that predicted by Hypothesis III.

B. The Fort Simpson Replication: Figure VI shows the means of the Fort Simpson Grade VII and VIII sample on the ten variables. The profile of the Fort Simpson Grade

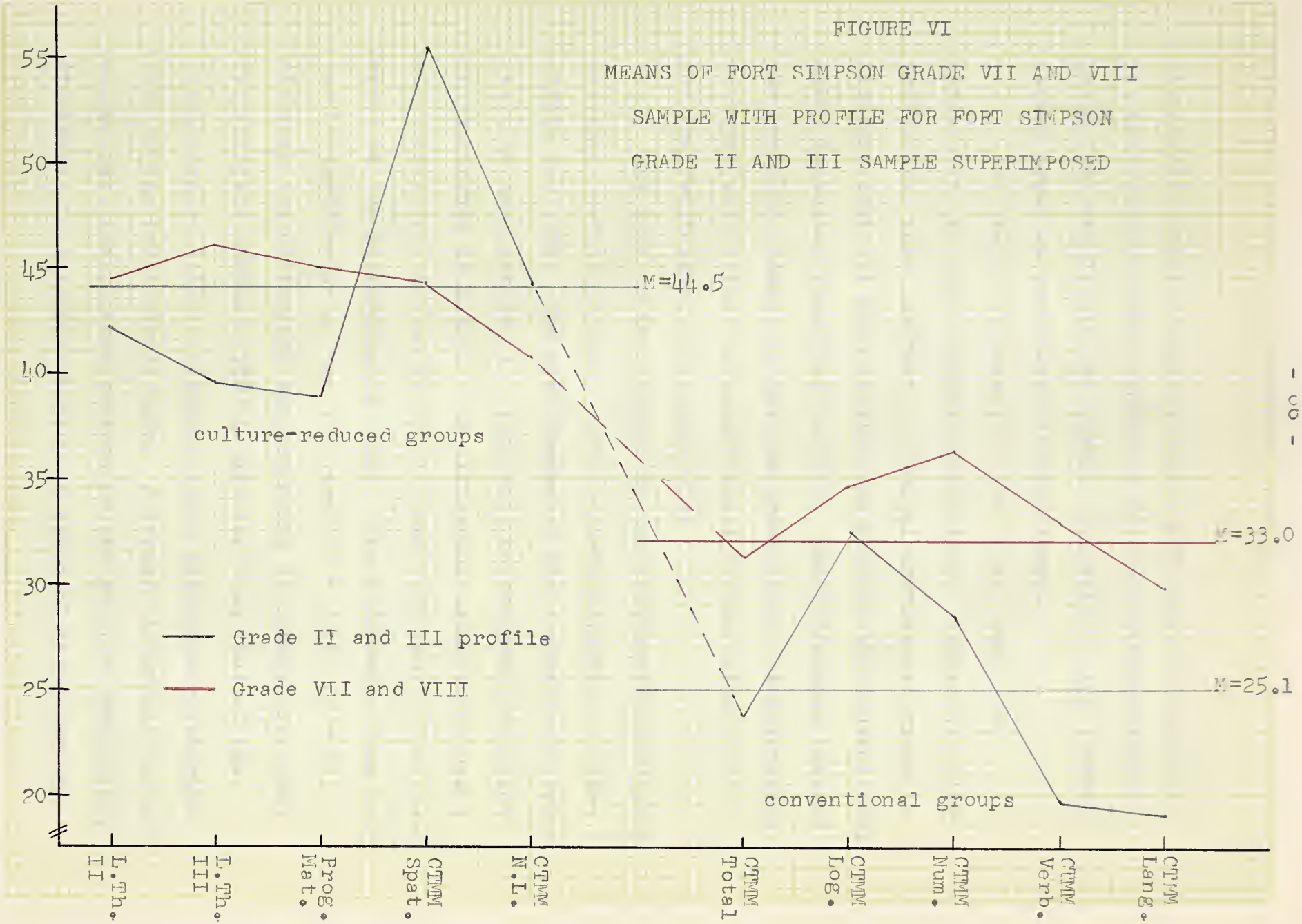


FIGURE VI

MEANS OF FORT SIMPSON GRADE VII AND VIII SAMPLE WITH PROFILE FOR FORT SIMPSON GRADE II AND III SAMPLE SUPERIMPOSED

II and III sample is superimposed upon that of the Grade VII and VIII such that the means of the culture-reduced group for both levels are equal. The mean of the conventional group at each level is also shown.

The trend discovered in the Faust study also appears in the Fort Simpson replication. Figure VI illustrates that when the mean of the culture-reduced group of tests is equal at both levels, the mean of the conventional group increases from 25.1 to 33.0. The difference between the means of changes in the two groups over the five-grade interval was tested for significance by the method used for the Faust sample (see page 83).

The mean of the changes in the culture-reduced group is again, obviously zero. The standard deviation of the change is 7.301. The mean change of the conventional group over the same period, is +7.88 and the standard deviation of the change is 1.777. The difference between the means of changes for the two groups is thus 7.88 and the standard error of the difference is 3.717. The difference does not reach significance at the .05 level ($t = 2.12$, $df = 8$). Although the difference between means of changes for the Fort Simpson sample (7.88) is greater than that of the Faust sample (5.29) it doesnot reach significance whereas that of the Faust sample does. A larger difference is required for significance because of the greater variability of changes in the culture-reduced group of tests for the

Fort Simpson sample. With but eight degrees of freedom the test of significance is not very powerful. Thus, although the results for the Fort Simpson replication only approach significance (P = .07), they nevertheless tend to support the findings of the Faust study.

Explanation of Results.

The Grade II and III profile was arbitrarily superimposed upon the Grade VII and VIII graph of means such that the mean of the culture-reduced group of means was the same at both levels. The shape of the Grade II and III profile, only, and not the level was assumed for the Grade VII and VIII class five grades earlier. Thus the cross-sectional data does not indicate whether there was a real increase or decrease in measured intelligence over the interval. The findings reported in this chapter, however, suggest that, either:

a. there is less decline in the conventional tests over a period of five grades than in the culture-reduced tests,

or

b. there is a greater increase in the conventional tests over a period of five grades than in the culture-reduced tests.

In other words, the conventional tests appear to be less biased against both samples at the Grade VII and VIII level than they are at the Grade II and III level. Since this result is exactly opposite to that predicted by Hypothesis

II some explanation is required.

There are several reasons why the results obtained may have occurred. Two of the most likely explanations are given.

1. Those students who do not adequately develop the verbal skills measured by conventional intelligence tests are likely to have difficulty in school and may be required to repeat several grades. By Grade VII and VIII, many pupils from the samples have dropped out of school. It is reasonable to assume that those pupils who score highest on the conventional and educationally loaded type of test would find school more profitable and therefore continue their education. As a result of such selectivity, the Grade VII and VIII conventional test scores may be expected to be relatively higher in a cross-sectional study.

2. The Faust and Fort Simpson schools may provide intellectually stimulating experience comparable to those provided for Edmonton and Calgary children and therefore tend to reduce the differential between the cultures. In accordance with postulate 8, the culture-reduced tests are expected to be less sensitive to the effects of environmental experiences and therefore would not show as great an increase as a result of the educational treatment received by the samples.

Hypothesis III has thus received no support from the data available for this investigation. Not only does a cross-sectional study prove inappropriate for testing the hypothesis, but schooling provided for culturally handicapped children may effectively reduce their handicap and thus act as an intervening variable altering the expected course of intellectual development. As a result of their educational opportunities the Metis and Indian children of Faust and Fort Simpson cannot be compared to the canal

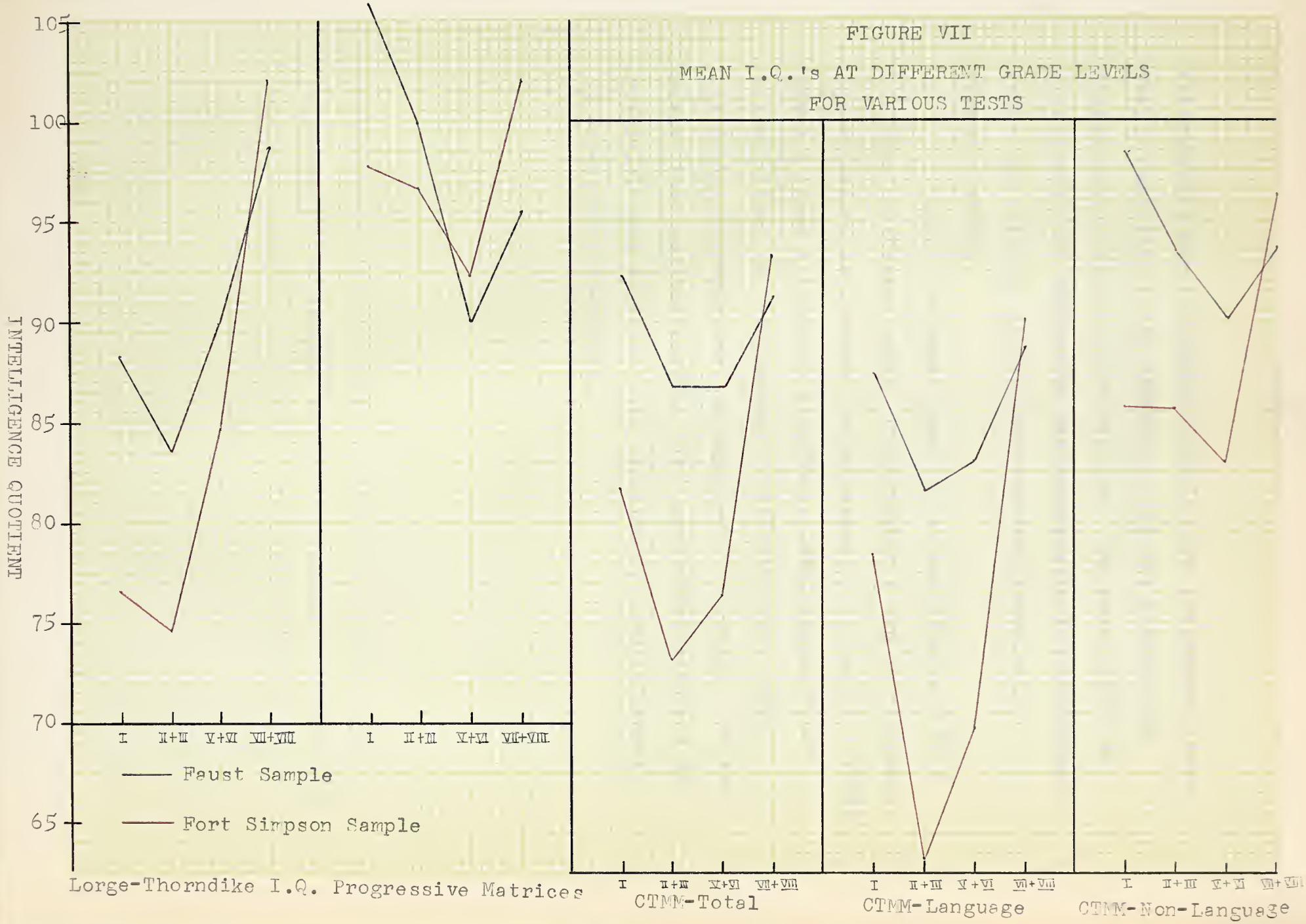
boat children of England or the children of isolated mountain communities in America previously mentioned.

The suggestion of selectivity resulting from academic failure and school drop outs about the Grade VII level is given support by Figure VII which shows the mean I.Q. of the different grade levels on several tests for both samples. Although the various tests are standardized on different norming groups, an effort is usually made to make the norming groups comparable for the different levels of a given test. No statistical significance is attached to the changes in I.Q. over the various grade levels, however it will be noted that the pattern of changes is remarkably similar for several tests and for both samples. The measured intelligence of the samples appears to decline during the elementary grades with respect to that of the norming population for the tests. It takes a sharp increase however at the junior high school level.

To explain the characteristic decline during the elementary grades and the eventual increase during the junior high school grades, a combination of Hypothesis III and the principle of selectivity seems most likely. However, it may be postulated that in a nation such as ours, with a reasonable standard of compulsory education for all, a relatively poor cultural environment will show its maximum effects on measured intelligence during the preadolescent years, and that with appropriate educational treatment, these

FIGURE VII

MEAN I.Q.'s AT DIFFERENT GRADE LEVELS FOR VARIOUS TESTS



deficiencies may be minimized during the adolescent years. Such a postulate is in agreement with the findings of Bradway and Robinson cited earlier. The possibility is also given some support by the increase in I.Q. of adolescents who attended the better secondary schools in Vernon's study.

It should be noted that it is intelligence B that has been considered here. Intelligence A and intelligence A', by definition, cannot be increased. As long as intelligence A' does not impose its limit, intelligence B, or present level of intellectual functioning may, under favorable circumstances, be expected to increase. To the extent that culture-reduced tests are better measures of potential they are also less affected by such influences as educational treatment.

CHAPTER VII

THE VALIDITY OF THE TESTS

Relationship of Tests with Other Tests of the Battery.

The concurrent validity of a test is usually demonstrated by computing the correlations of the test with other well known measures of intelligence. To provide some indication of how well each of the tests correlate with other variables of the battery at a given level, a correlation matrix of each battery is given in Tables XVII to XXIV. In these tables the column headed "Mean" gives the average of all non-spurious correlations of the test with every other variable of the battery. Correlations were averaged by the method suggested by Garrett (1958), page 173. In this method the r's are transformed into Fisher's z function and the arithmetic mean of the z's is found. The resulting mean z is then converted into an equivalent r.

A low correlation between Test X and Test Y does not indicate that Test X lacks validity for a desired purpose unless it can be assumed that Test Y validly measures that which is desired. Similarly, a low mean correlation may result from the lack of validity of other tests in the battery. Since several of the sub-tests are rather short, their reliability and consequently their validity is likely to be somewhat limited. Tables XVII to XXIV must

therefore be interpreted with caution. It should also be noted that with fairly small N's, rather large correlation coefficients are required for significance at the .05 level.

TABLE XVII

CORRELATION MATRIX AND MEAN OF NON-SPURIOUS CORRELATIONS FOR THE FAUST GRADE I BATTERY N = 32

	2	3	4	5	6	7	Mean
1. L. Th. 1	26	47	72	27	24	10	.27
2. L. Th. II		46	73	52	41	22	.38
3. L. Th. III			85	33	38	07	.35
4. L. Th. Total				47	45	16	.36
5. Prog. Mat.					34	08	.35
6. S C R I T						03	.31
7. Detroit							.11

Decimal points are omitted from matrix.

All correlation of .35 or higher are significant at the .05 level.

TABLE XVIII

CORRELATION MATRIX AND MEAN OF NON-SPURIOUS CORRELATIONS FOR FORT SIMPSON GRADE I BATTERY N = 19

	2	3	4	5	6	7	Mean
1. L. Th. I	23	53	73	18	39	44	.36
2. L. Th. II		29	70	50	50	50	.41
3. L. Th. III			75	41	19	54	.40
4. L. Th. Total				42	41	63	.49
5. Prog. Mat.					64	56	.46
6. S C R I T						46	.45
7. Detroit							.52

Decimal points are omitted from matrix.

All correlations of .46 or higher are significant at the .05 level.

TABLE XIX

CORRELATION MATRIX AND MEAN OF NON-SPURIOUS CORRELATIONS FOR THE FORT SIMPSON GRADE II AND III BATTERY N = 46

	2	3	4	5	6	7	8	9	10	11	12	13	Mean
1. L. Th. I	22	-07	41	-17	-29	10	09	00	-15	34	10	07	.02
2. L. Th. II		52	78	20	28	56	39	48	29	51	48	46	.40
3. L. Th. III			73	37	59	42	18	33	25	42	36	27	.34
4. L. Th. Total				17	32	56	29	42	24	65	52	36	.40
5. Prog. Mat.					51	15	22	21	09	00	03	29	.18
6. S C R I T						41	34	38	46	12	22	52	.34
7. CTMM Total							48	78	73	74	88	73	.38
8. CTMM-Spatial								23	25	17	13	85	.23
9. CTMM-Logic									55	51	79	54	.35
10. CTMM-Num										30	62	56	.26
11. CTMM-Verb.											83	34	.35
12. CTMM-Lang.												39	.29
13. CTMM-N.L.													.34

Decimal points are omitted from matrix.

All correlations of .29 or higher are significant at the .05 level.

TABLE XX

CORRELATION MATRIX AND MEAN OF NON-SPURIOUS CORRELATIONS FOR THE FAUST GRADE II AND III BATTERY N = 42

	2	3	4	5	6	7	8	9	10	11	12	13	Mean
1. L. Th. I	27	06	55	36	24	42	34	34	27	45	39	38	.32
2. L. Th. II		38	79	54	68	55	39	27	41	62	52	49	.48
3. L. Th. III			72	23	17	23	22	18	12	20	26	21	.21
4. L. Th. Total				56	54	59	46	38	42	60	57	53	.52
5. Prog. Mat.					79	75	50	46	62	63	68	63	.58
6. S C R I T						65	44	44	47	60	57	54	.52
7. CTMM Total							73	74	78	73	85	87	.55
8. CTMM-Spatial								48	36	49	40	91	.41
9. CTMM-Logic.									52	33	70	59	.38
10. CTMM-Num.										50	76	64	.42
11. CTMM-Verb.											75	58	.51
12. CTMM-Lang.												55	.50
13. CTMM- N.L.													.50

Decimal points are omitted from matrix.

All correlations of .30 or higher are significant at the .05 level.

TABLE XXI

CORRELATION MATRIX AND MEAN OF NON-SPURIOUS CORRELATIONS FOR THE FAUST GRADE V AND VI BATTERY N = 29

	2	3	4	5	6	7	Mean
1. L. Th. I	47	44	79	64	48	73	.57
2. L. Th. II		45	75	62	53	56	.53
3. L. Th. III			62	69	56	38	.51
4. L. Th. Total				81	64	67	.72
5. Prog. Mat.					72	73	.71
6. Cattell						75	.62
7. S C R I T							.65

Decimal points are omitted from matrix.

All correlations of .37 or higher are significant at the .05 level.

TABLE XXII

CORRELATION MATRIX AND MEAN OF NON-SPURIOUS CORRELATIONS FOR THE FORT SIMPSON GRADE V AND VI BATTERY N = 58

	2	3	4	5	6	7	Mean
1. L. Th. I	62	56	76	45	48	35	.50
2. L. Th. II		77	91	59	65	49	.63
3. L. Th. III			91	65	64	43	.62
4. L. Th. Total				67	67	47	.61
5. Prog. Mat.					41	35	.53
6. Cattell						46	.56
7. S C R I T							.43

Decimal points are omitted from matrix

All correlations of .26 or higher are significant at the .05 level.

TABLE XXIII

CORRELATION MATRIX AND MEAN OF NON-SPURIOUS CORRELATIONS FOR THE FAUST GRADE VII AND VIII BATTERY N = 23

	2	3	4	5	6	7	8	9	10	11	12	13	Mean
1. L. Th. I	27	66	86	54	40	34	51	34	44	49	37	57	.45
2. L. Th. II		30	47	29	27	12	25	52	12	20	29	26	.25
3. L. Th. III			89	60	48	56	61	55	58	63	70	76	.59
4. L. Th. Total				63	50	54	62	56	55	61	66	75	.60
5. Prog. Mat.					56	39	69	66	47	61	52	69	.56
6. Cattell						37	51	40	52	55	45	61	.47
7. CTMM-Spatial							30	17	17	18	88	48	.32
8. CTMM-Logic.								59	58	77	38	77	.53
9. CTMM-Num.									47	65	52	71	.48
10. CTMM-Verb.										94	36	88	.45
11. CTMM-Lang.											38	93	.47
12. CTMM-N.L.												66	.48
13. CTMM-Total													.63

Decimal points are omitted from matrix.

All correlations of .41 or higher are significant at the .05 level.

TABLE XXIV

CORRELATION MATRIX AND MEAN OF NON-SPURIOUS CORRELATIONS FOR THE FORT SIMPSON GRADE VII AND VIII BATTERY N = 32

	2	3	4	5	6	7	8	9	10	11	12	13	Mean
1. L. Th. I	32	48	48	11	17	20	41	45	61	63	19	58	.39
2. L. Th. II		46	70	57	67	13	25	31	03	12	23	21	.32
3. L. Th. III			84	47	44	12	46	41	44	51	21	52	.41
4. L. Th. Total				51	54	19	50	55	50	59	27	60	.48
5. Prog. Mat.					64	15	30	25	19	24	18	29	.34
6. Cattell						11	33	22	01	12	13	16	.32
7. CTMM-Spatial							31	17	06	10	87	41	.16
8. CTMM-Logic.								59	57	72	36	73	.42
9. CTMM-Num.									54	71	35	74	.40
10. CTMM-Verb.										96	07	86	.32
11. CTMM-Lang.											12	91	.33
12. CTMM-N.L.												48	.18
13. CTMM-Total													.40

Decimal points are omitted from matrix.

All correlations of .36 or higher are significant at the .01 level.

Relationship of Tests With Achievement.

Since Binet published the first practical intelligence test, correlation with school achievement has been used as a principal criterion for the validation of measures of intelligence.

A test which is valid for one group may not be valid when it is used with subjects who fall outside a specific cultural or educational range. Validity depends on the characteristics of the group to which the test is applied. It is therefore necessary to establish the validity of the tests for the samples involved in this study.

Chapter II suggests that a very high correlation with achievement is undesirable for cross-culture tests or for measures from which we wish to estimate undeveloped potential. Very high correlations with achievement indicate a high educational loading. A moderate correlation is necessary, however, in order to indicate that the test does sample from those abilities required for academic success.

Table XXV gives the correlation of each test with the California Achievement Total grade placement for both samples at each level. It will be noted that with few exceptions the correlations are significant and a great majority are substantial (.40 or more).

TABLE XXV

CORRELATION OF VARIOUS TESTS WITH THE CALIFORNIA ACHIEVEMENT TOTAL GRADE PLACEMENT AT EACH LEVEL

	GRADE I		GRADE II + III		GRADE V + VI		GRADE VII + VIII	
	Faust N=32	Ft.Spn. N=19	Faust N=42	Ft.Spn. N=46	Faust N=29	Ft.Spn. N=58	Faust N=23	Ft.Spn. N=32
Lorge-Thorndike I	.35*	.47*	.54**	.02	.53**	.37**	.48*	.58**
Lorge-Thorndike II	.63**	.15	.55**	.58**	.64**	.65**	.19	.43*
Lorge-Thorndike III	.47**	.49*	.30*	.63**	.52**	.50**	.60**	.71**
Lorge-Thorndike Total	.62**	.43	.65**	.64**	.73**	.60**	.63**	.75**
Ravens Prog. Matrices	.51**	.55*	.72**	.30*	.62**	.24	.58**	.43*
SCRIT	.53**	.44	.64**	.47**	.51**	.47**		
Cattell					.45*	.41**	.56**	.38*
CTMM - Spatial			.43**	.30*			.31	.27
- Logical			.60**	.49**			.71**	.57**
- Numerical			.69**	.35*			.56**	.59**
- Verbal			.74**	.49**			.74**	.71**
- Language			.77**	.56**			.80**	.75**
- Non-Language			.60**	.41**			.47*	.31
- Total			.77**	.56**			.80**	.77**
Detroit Beginning	.28	.78**						
Mean for Battery	.49**	.49**	.63**	.47**	.56**	.47**	.60**	.59**

* Significant at .05 level
** Significant at .01 level

CHAPTER VIII

EVALUATION OF THE TESTS

Criteria for Evaluation.

Several criteria for the evaluation of a cross-culture test of intellectual potential were set out in Postulate 9, page 28. An adequate assessment of the tests in the experimental battery at each level should be based upon each and every criterion stated in the postulate. The present investigation, however, does not permit such a complete evaluation. Emperical evidence necessary for the use of criteria (a) and (b) was not obtained in this investigation but will be provided by the factor analytic study carried out by Dr. MacArthur. Furthermore, the cross-sectional data proved to be inadequate for the assessment of tests according to criterion (d). In spite of these limitations, the findings of the present study do permit a partial evaluation of the tests against four criteria of considerable importance. These criteria are listed below.

1. A cross-culture test of intellectual potential should show less difference between cultures in a bi-cultural administration than do conventional verbal intelligence tests.

2. It should contain items that can be solved in any language or mode of expression and which are likely to be as familiar and useful for one cultural group as another.

3. It should show substantial relationship to school achievement.

4. It should show a significant relationship with other well known and commonly used measures of intelligence.

Criterion 4, above, is used to assure that the tests have some measure of concurrent validity for the samples tested. It is a tentative substitute for criteria (a) and (b) of Postulate 9.

Appraisal of the Tests.

The results from testing Hypothesis I to IV suggest that the degree to which a given test meets the evaluation criteria is a function of the level at which the test is administered. In other words, a given test may satisfy the criteria at one level but may fail to do so at another. For this reason it is necessary to evaluate the tests separately for each grade level.

As the major purpose of this investigation is to identify those tests which show a minimum of cultural bias for the two samples studied, not all criteria are considered of equal importance. The tests will be evaluated, therefore, primarily against criterion 1 of the preceding section, provided that they meet the minimum requirements of criteria 2, 3, and 4. A test shall be regarded as having met criterion 2 if it has been classified a priori as culture-reduced. See page 29. It shall be regarded as

having met criterion 3 if it shows a substantial correlation (.40 or more) withthe California Achievement Total grade placement, provided however, that all such correlations are significant at the .05 level. Finally, a test shall be regarded as having met criterion 4 if it shows a mean non-spurious correlation with other tests of the battery which is significant at the .05 level.

Tables XXVI to XXIX summarize the major findings of the investigation and show how each test rates against the four selected criteria.

The experimental findings about a test do not always coincide for both the Faust and Fort Simpson samples. To illustrate this fact it will be noted that differences which are significant for one sample may or may not be significant for the other. Although not identical the findings for both samples are remarkably similar.

It is possible that the two samples are representative of entirely different populations and that some tests have greater cultural bias for one population than for the other. However, if we risk to generalize about a population which includes both samples we may get a more realistic evaluation of the tests for the cross-culture assessment of the intellectual potential of Canadian rural, non-white, or native groups. The results for both samples are combined, therefore, in the evaluation tables which follow. The combination of results has been achieved by finding the

weighted mean of the results for the two samples.

The weights used for finding the combined σ-deviations were the N's for each sample. It can be shown that the results thus obtained are identical to those which would be produced if the T-scores for both samples were first pooled, and the mean of the combined group then found. The column headed σ-deviation therefore gives the deviation of the mean of the combined group from the Edmonton or Calgary T-score mean of 50 in units of the standard deviation of the norming distribution. As before, the σ-deviation is interpreted as an indication of the extent of cultural bias in the test. The tests are listed in Tables XXVI to XXIX according to increasing size of their σ-deviations.

Fisher's z transformations were used for computing the combined correlations. The weights used for this purpose were the N's - 3 for both samples. See Garrett (1958), page 173. The degree of freedom used for testing the significance of the combined r's was $(N_1 - 3) + (N_2 - 3) - 2$.

Columns headed "r_A" give the correlation of the tests with the California Achievement Total grade placement. Columns headed "r_0" give the mean non-spurious correlation of the tests with all other tests of the battery.

Columns headed "Faust: significantly less biased than -" and "Ft. Spn.: significantly less biased than -" give the identification numbers of those tests which show

significantly greater bias than the test under consideration for both the Faust and Fort Simpson samples respectively.

Columns headed II, III, and IV tell whether the test under consideration meets the requirements of criteria 2, 3, and 4 respectively.

TABLE XXVI

SUMMARY OF FINDINGS FOR THE GRADE I BATTERY OF TESTS AS EVALUATED AGAINST FOUR CRITERIA

No.	Test	σ-dev.	Faust: Significantly less biased than --	Ft.Spn.: Significantly less biased than --	*	r_A	r_O	II	III	IV
1.	S C R I T	.41	2,3,4,5,6,7,	6,7,	R	.50	.36	Yes	Yes	Yes
2.	L. Th. II	.76	5,6,7,	6,7,	R	.49	.39	Yes	Yes	Yes
3.	L. Th. III	.81	5,6,7,	5,6,7,	R	.48	.37	Yes	Yes	Yes
4.	Prog. Mat.	.89	7,	7,	R	.52	.39	Yes	Yes	Yes
5.	L. Th. Total	1.24			C	.56	.41	No	Yes	Yes
6.	L. Th. I	1.44			C	.40	.30*	No	Yes	No
7.	Detroit Beg.	1.78			C	.51	.27*	No	Yes	No

* - Classification: C - conventional: R - culture-reduced.
r_A - correlation with achievement
r_O - mean correlation with other tests of the battery.
* - not significant at .05 level.

TABLE XXVII

SUMMARY OF FINDINGS FOR THE GRADE II AND III BATTERY OF TESTS AS EVALUATED AGAINST FOUR CRITERIA

No.	Test	σ-dev.	Faust: Significantly less biased than --	Ft.Spn.: Significantly less biased than --	**	r_A	r_O	II	III	IV
1.	CTMM-Spat.	.00	2,4,5,6,7,8,9,10,11, 12,13,	2,3,4,5,6,7,8,9,10,11 12,13,	R	.45	.32	Yes	Yes	Yes
2.	CTMM-N.L.	.87	6,7,8,9,10,11,12,13,	6,7,8,9,10,11,12,13,	R	.51	.42	Yes	Yes	Yes
3.	L. Th. III	.90	5,6,7,8,9,10,11,12 13,	7,8,9,10,11,12,13,	R	.49	.28	Yes	Yes	Yes
4.	L. Th. II	.95	5,6,7,8,9,10,11,12, 13,	7,8,9,10,11,12,13,	R	.56	.44	Yes	Yes	Yes
5.	S C R I T	1.09	6,8,9,10,11,12,13,	7,8,9,10,11,12,13,	R	.56	.43	Yes	Yes	Yes
6.	Prog. Mat.	1.43	10,11,12,13,	7,8,9,10,11,12,13,	R	.54	.39	Yes	Yes	Yes
7.	CTMM-Log.	1.76	10,11,12,13,	8,9,10,11,12,13,	C	.55	.37	No	Yes	Yes
8.	CTMM-Num.	2.12	13,	10,11,12,13,	C	.54	.34	No	Yes	Yes
9.	L. Th. Total	2.16	11,12,13,	12,13,	C	.64	.46	No	Yes	Yes
10.	CTMM-Total	2.50	13,	12,13,	C	.67	.47	No	Yes	Yes
11.	L. Th. I	2.53	13,	13,	C	.29	.17*	No	No	No
12.	CTMM-Verb.	2.75			C	.63	.39	No	Yes	Yes
13.	CTMM-Lang.	2.92			C	.67	.40	No	Yes	Yes

** Classification: C - conventional. R - culture-reduced
r_A - correlation with achievement
r_O - mean correlation with other tests of the battery
* - not significant at .05 level.

TABLE XXVIII

SUMMARY OF FINDINGS FOR THE GRADE V AND VI BATTERY OF TESTS AS EVALUATED AGAINST FOUR CRITERIA

No.	Test	σ-dev.	Faust: Significantly less biased than --	Ft. Spn.: Significantly less biased than --	*	r_A	r_O	II	III	IV
1.	Prog. Mat.	1.10	7,	4,6,5,7,	R	.38	.50	Yes	No	Yes
2.	Cattell	1.15	7,	4,6,5,7,	R	.42	.51	Yes	Yes	Yes
3.	S C R I T	1.19	7,	4,6,5,7,	R	.48	.44	Yes	Yes	Yes
4.	L. Th. I	1.50	7,	5,6,7,	R	.42	.44	Yes	Yes	Yes
5.	L. Th. III	1.92			R	.63	.48	Yes	Yes	Yes
6.	L. Th. II	1.92			R	.65	.48	Yes	Yes	Yes
7.	L. Th. Total	2.19			R	.60	.49	Yes	Yes	Yes

* - Classification: C - conventional: R - culture-reduced
** r_A - correlation with achievement
r_O - mean correlation with other tests of the battery

TABLE XXIX

SUMMARY OF FINDINGS FOR THE GRADE VII AND VIII BATTERY OF TESTS
AS EVALUATED AGAINST FOUR CRITERIA

No.	Test	σ-dev.	Faust: Significantly less biased than --	Ft.Spn.: Significantly less biased than --	**	r_A	r_O	II	III	IV
1.	L. Th. I	.44	5,6,9,10,11,12,13,	9,10,11,12,13,	R	.54	.42	Yes	Yes	Yes
2.	CTMM-Spat.	.45	9,10,11,12,13,	9,10,11,12,13,	R	.29	.23*	Yes	No	No
3.	L. Th. III	.52	11,12,13,	9,10,11,12,13,	R	.67	.49	Yes	Yes	Yes
4.	L. Th. Total	.67	11,12,13,	9,10,11,12,13,	R	.71	.54	Yes	Yes	Yes
5.	Prog. Mat.	.67	13,	9,10,11,12,13,	R	.49	.44	Yes	Yes	Yes
6.	L. Th. II	.69	13,	9,10,11,12,13,	R	.34	.29	Yes	No	Yes
7.	CTMM-N.L.	.80	9,10,11,12,13,	10,11,12,13,	R	.49	.31	Yes	Yes	Yes
8.	Cattell	.89	13,	10,11,12,13,	R	.45	.38	Yes	Yes	Yes
9.	CTMM-Num.	1.34		13,	C	.58	.43	No	Yes	Yes
10.	CTMM-Log.	1.38			C	.63	.47	No	Yes	Yes
11.	CTMM-Verb.	1.59			C	.72	.37	No	Yes	Yes
12.	CTMM-Total	1.67			C	.78	.51	No	Yes	Yes
13.	CTMM-Lang.	1.91			C	.77	.39	No	Yes	Yes

** Classification: C - conventional: R - culture-reduced
r_A - correlation with achievement
r_O - mean correlation with other tests of the battery
* - not significant at .05 level.

A. The Grade I Battery: From Table XXVI it will be seen that the SCRIT, Lorge-Thorndike II (Cross-Out: the one that does not belong), Lorge-Thorndike III (Pairing: the two that go together), and Progressive Matrices each satisfy all four criteria. All are significantly less biased than the Detroit Beginning First Grade Intelligence Test. Differences in extent of cultural bias among these tests are small; however, the SCRIT probably shows greatest promise as a cross-culture test at this level. It has least bias for both samples. For the Faust sample it shows significantly less bias than any other test.

B. The Grade II and III Battery: Table XXVII shows that each of the following tests meet all of the four criteria.

- CTMM Spatial
- CTMM Non-Language
- Lorge-Thorndike II (Cross-Out: the one that does not belong)
- Lorge-Thorndike III (Pairing: the two that go together)
- S C R I T
- Progressive Matrices

The CTMM Spatial and the Lorge-Thorndike III have rather marginal correlations with other tests of the battery. This may cast some doubt on their validity. It will be noted that for the combined group the CTMM Spatial shows no cultural bias at this level. Indeed the combined sample did somewhat better on this test than their Edmonton and Calgary counterparts. The CTMM Non-Language test probably shows

greatest promise as a cross-culture test at this level. It is fairly long and highly reliable. For both samples it shows significantly less bias than the Progressive Matrices and all conventional tests of the battery. It correlates substantially both with achievement and with the other tests of the battery. Lorge-Thorndike II also shows exceptional promise.

C. The Grade V and VI Battery: There are no conventional tests in the experimental battery at this level. Hence, all tests are assumed to have met criteria 1 and 2. The Progressive Matrices falls somewhat short of the standard set for criterion 3. Garrett (1958), page 176, defines a "substantial correlation" as one of .40 or higher. Hence, this was the standard set for criterion 3. The Progressive Matrices correlates only .38 with achievement at this level. The test, however, shows least cultural bias and has a comparatively high correlation with the other tests of the battery. The low correlation with achievement resulted from the contribution made by the Fort Simpson Sample. See Table XXV. The Grade V and VI group of the Fort Simpson sample contained several special classes with diversified rather than standard curricula. Students from special classes regardless of their intelligence may be unprepared for a standardized achievement test, and hence lower the correlation. The test correlates .62 with achievement for the Faust sample.

Further investigation at this level is required before it should be assumed that the Progressive Matrices does not meet the criteria of a cross-culture test of intellectual potential.

The Lorge-Thorndike test at this level shows significantly greater cultural bias toward the samples than the Progressive Matrices, Cattell or SCRIT. Indeed the extent of cultural bias shown by the Lorge-Thorndike would suggest that it has little promise as a cross-culture test.

D. The Grade VII and VIII Battery: Six tests at this level meet all four criteria.

- Lorge-Thorndike I (Figure Classification)
- Lorge-Thorndike III (Figure Analogies)
- Lorge-Thorndike Total
- Progressive Matrices
- CTMM Non-Language
- Cattell

The Lorge-Thorndike Total is exceptionally promising at this level as it correlates highly with achievement and has the highest correlation with the other tests of the battery. Such evidence points toward a high validity. The test is also highly reliable. It shows no greater cultural bias than the Progressive Matrices and is significantly less biased toward both samples than nearly all conventional tests of the battery. Lorge-Thorndike I and III are in themselves promising tests. Lorge-Thorndike II (Number Series) has a low correlation with achievement and also a low correlation with the other tests of the battery.

A combination of Lorge-Thorndike I and III may therefore make a better cross-culture test than the Lorge-Thorndike Total.

The Cattell, Progressive Matrices, and CTMM Non-Language also show superior promise as cross-culture tests of potential at this level.

CHAPTER IX

SUMMARY, CONCLUSIONS, AND IMPLICATIONS

The purpose of this investigation was to study a selection of promising culture-reduced intelligence tests, and to identify those which for two samples of Metis and Indian children at four grade levels show a minimum of cultural bias.

With the possible exception of SCRIT each test in the experimental battery at the four levels is well known and commonly used in psychology and education for the assessment of intellectual ability. A review of the evidence on the reliability and validity of the tests was made. The problem of ascertaining the construct validity of the tests as measures of intellectual potential was left to the factor analytic study by Dr. R.S. MacArthur. In order to assess the extent of cultural bias, raw scores were converted to derived scores based on a scale of T-scores for Edmonton and Calgary students. Such T-scores have a normal distribution, a mean of 50, and a standard deviation of 10. The sample mean on each test was then compared to the Edmonton or Calgary mean of 50. Those tests for which the sample mean deviated least from the Edmonton or Calgary mean were considered to show least cultural bias, or least discrimination in a bi-cultural administration.

Differences between sample means on the various

tests of the battery were then tested for statistical significance. Four hypotheses were tested for each sample. Finally, the results for both samples were combined and each test evaluated against four criteria. The evaluation of the tests for cross-culture measurement of intellectual potential is admittedly partial and tentative. The major consideration in the evaluation was the extent of cultural bias shown by the test. Two other criteria -- a) a high loading on the general factor, and b) negligable group factor loadings -- are deemed to be important by the theoretical framework upon which the study is based. These criteria are not employed in the present appraisal however, as the necessary data were not available at the time of writing. A substitute criterion is used to assure that the tests are measuring some sort of common factor for the samples studied. When completed the findings of Dr. MacArthur's factor analysis of the tests will permit the application of the two omitted criteria by the interested reader.

Experimental Findings

1. For both the Faust and Fort Simpson samples, and at each of the four grade levels, some of the tests of the experimental battery were found to show significantly less cultural bias than others. These tests instruments have been identified. Hypothesis I has received support,

therefore, from all data of the study. The concept of cultural bias in tests of intelligence appears to be well founded.

2. In each instance where it was possible to test Hypothesis II it was found that tests which:

(a) consist largely of items that can be solved in any language or mode of expression,

(b) have minimal dependence on past specific learning, and

(c) are probably as novel to one culture group as to another,

showed significantly less cultural bias than other more conventional tests.

3. Although the cross-sectional data proved to be inadequate for testing Hypothesis III, it was found that the culture-reduced tests show significantly less increase with grade level, probably as a result of school treatment and selection, than do traditional educationally loaded tests.

4. Where N is not too small (considerably less than 30) those tests which show a minimum of cultural bias also show, with few exceptions, significant and substantial correlations with academic achievement. Although having a minimum dependence on past specific learning, it would appear that these tests do sample from the abilities required for academic success.

5. With the advent of factor analytic techniques

it has become known that tests bearing the same name may measure somewhat different intellectual functions at different age levels. An item at one level may assess reasoning ability whereas the same item at a different level may be merely a test of memory. Analogous to changes in the factorial content of a test from one level to another are changes in the extent of cultural bias it possesses. A test which appears to show little cultural bias at one level may show considerably more bias at another level. The Lorge-Thorndike, for example, shows little bias at the Grade VII and VIII level but a large amount at the Grade V and VI level.*

Conclusions.

In Chapter II it was postulated that tests heavily loaded with the statistical factor g and having minimum cultural bias provide our best estimates of present potential or intelligence A'. The present investigation has identified those tests which for two samples of Canadian Metis and Indian children show a minimum of cultural bias.

---

* At the Grade VII and VIII level, the Lorge-Thorndike showed no greater bias than the Progressive Matrices, (σ-deviation .67 for both tests). At the Grade V and VI level, however, the Lorge-Thorndike combined sample mean deviated twice as far from the Calgary mean of 50 as did the sample mean for the Progressive Matrices (σ-deviation were 2.19 and 1.10 respectively).

Of the tests investigated, those which satisfy the four criteria have been ranked at each level according to the absence of cultural bias shown. Assuming the postulate stated above, we may conclude that for the most adequate assessment of Intelligence A', test instruments should be chosen from those listed below according to:

(a) their g loading,

(b) absence of group factors, and

(c) rank in list.

Grade I Level.

1. SCRIT
2. Lorge-Thorndike II and Lorge-Thorndike III
3. Progressive Matrices

Grade II and III Level.

1. CTMM Non-Language
2. Lorge-Thorndike II and Lorge-Thorndike III
3. SCRIT
4. Progressive Matrices.

Grade V and VI Level.

1. Progressive Matrices
2. Cattell
3. SCRIT

Grade VII and VIII Level.

1. Lorge-Thorndike I and II or Total
2. Progressive Matrices
3. CTMM Non-Language
4. Cattell

It will be noted that the Progressive Matrices is recommended at all levels. The SCRIT also meets the four criteria at the three levels for which it is appropriate.

The CTMM Non-Language test satisfies the criteria at both levels for which T-score norms were available.

In the absence of information regarding the g loadings and the presence of group factor loadings, each test listed may be considered a promising instrument for the cross-culture assessment of intelligence. Each may be used with assurance that it is measuring, at least to some degree, a sample of the intellectual abilities that are measured by other well known tests and that are related to academic success. Furthermore, each test instrument listed may also be counted upon to measure these abilities with significantly less cultural bias than do conventional tests.

Implications.

1. Theoretical. The theory advanced at the outset of this study has received some support from the experimental findings. The verification of three hypotheses adds credence to the postulates from which they are derived. Specifically, different cultures provide differential opportunities to learn information, skills, attitudes, work habits, etc. required for successful performance on an intelligence test. Some tests are significantly less affected by such cultural differences than are others. Tests or sub-tests which depend upon verbal symbolism and acquired information differentiate more clearly between cultures than do non-verbal reasoning tests. We know that

it is therefore possible to design tests for cross-culture use so as to minimize the effects of differential opportunities to learn specific skills or acquire specific information. Furthermore, the study adds to our knowledge and theory regarding the types of tests or sub-tests (ie. the types of skills and information) which differentiate most clearly between cultures.

2. Practical. The tests identified as showing a minimum of cultural bias in a bi-cultural administration are likely to have value for the following purposes:

(a) identification of superior pupils from underprivileged backgrounds.

(b) making improved cross-cultural comparisons.

(c) distinguishing between mental deficiency and language or reading difficulties.

(d) identification of underachievement.

(e) predicting to long term and nonspecific criteria when treatment is adapted to ability.

(f) assessing the intellectual ability of those who have little or no formal education.

3. Implications for Further Research. Longitudinal cross-culture studies are needed which begin early in childhood and continue well into adulthood. Such studies would provide some indication of the pattern of intellectual development of underprivileged groups from minority cultures. The experimental subjectsof such studies could also be introduced to novel treatments at different age levels in order to determine:

(a) those factors which stimulate intellectual development

and

(b) whether there are, in fact, critical periods in the development of intellectual abilities.

The small numbers of Metis and Indian children, at a given age level, who are available for testing in the same community also argue in favor of a more intensive longitudinal study. Moreover, this study points to the need for longitudinal data in order to test Hypothesis III.

The tests should also be subjected to intensive item analysis and a recombination of sub-tests or items should be attempted in order to maximize reliability and validity as well as to minimize cultural dependence.

B I B L I O G R A P H Y

BIBLIOGRAPHY

Anastasi, Anne. Psychological Testing. (2nd ed.), New York: Macmillan Company, 1961.

Anastasi, Anne. Differential Psychology. New York: Macmillan Company, 1958.

Bayley, Nancy. Proceedings, 1956 Invitational Conference on Testing Problems. Princeton: E.T.S., 1957.

Bayley, N., "Factors Influencing the Growth of Intelligence in Young Children". Year Bk. N.S.S.E., 39: 49-79, 1940.

Bayley, N. and Jones H.E., "Environmental Correlates of Mental and Motor Development: A cumulative Study From Infancy to Six Years". Child Devel. 8: 329-341, 1937.

Bingham, W.E. and Griffiths W.J., "The Effects of Different Environments During Infancy on Adult Behavior in the Rat". J. Comp. Physiol. Psych., 45: 307-312, 1952.

Bradway, K.P. and Robinson, N.M., "Significant I.Q. Changes in Seventy-Five Years: A Follow-Up". J. of Ed. Psych. Vol. 52: 2, 74-79, 1961.

Burke, H.R., "Raven's Progressive Matrices: Review and Critical Evaluation". J. Genet. Psychol. 93: 199-228, 1958.

Buros, O.K., (Edition) Fifth Mental Measurement Year Book. New Jersey: Gryphon Press, 1958.

California Achievement Tests: Primary. Los Angeles: California Test Bureau, 1951.

California Achievement Tests: Junior High Bevel. Los Angeles: California Test Bureau, 1957.

California Short-Form Test of Mental Maturity: Primary, Los Angeles: California Test Bureau, 1953.

California Short-Form Test of Mental Maturity: Elementary, Los Angeles: California Test Bureau, 1957.

Cattell, R.B., Handbook of Individual or Group Culture Free Intelligence Test, Scale 2. Champaign: IPAT, 1958.

Clark, R.S., Heron, W., Fehterstonbaugh, M.L., Forgays, D.O., "Individual Differences in Dogs: Preliminary Report on the Effects of Early Experience". Can. J. Psych., 5: 150-156, 1951.

Clarke, S.C.T., "Review of Some Previous Studies on Matriculation Problems". The Alta. J. of Ed. Res. 4: 1, 30-39, 1958.

Cronbach, L.J., Essentials of Psychological Testing. New York: Harpers, 1960.

Cronbach, L.J., and Glesser, G.C., Psychological Tests and Personnel Decisions. Urbana: University of Illinois Press, 1957.

Dennis, W., "Spalding's Experiment on the Flight of Birds Repeated with Another Species". J. Comp. Psych. 31: 337-348, 1941.

Detroit Beginning First-Grade Intelligence Test, (Revised). Manual of Directions and Key. New York: World Book Company, 1937.

Edwards, A.L., Experimental Design in Psychological Research. New York: Holt, Rinehart and Winston, 1950.

Eells, K.W., Intelligence and Cultural Differences. Chicago: University of Chicago Press, 1951.

Elley, W.B., "A Comparative Analysis of Socio-Economic Bias in Selected Intelligence Tests". Unpublished Ph. D. Thesis, University of Alberta, 1961.

Forgays, D.G. and Forgays, J.W., "The Nature of the Effects of Free-Environmental Experience on the Rat". J. Comp. Physiol. Psych. 45: 322-328, 1952.

Garrett, H.E., Statistics in Psychology and Education. New York: Logmans, Green and Company, 1958.

Garrett, H.E., "A Developmental Theory of Intelligence". American Psychologist. 1: 372-378, 1946.

Goldfarb, W., "Emotional and Intellectual Consequences of Psychologic Deprivation in Infancy: A Revaluation", in Hoch, P.H., and Zubin, J., Psychopathology of Childhood. New York: Grune and Stratton, 1955.

Good, W.R., "Misconceptions About Intelligence Testing", The University of Michigan School of Education Bulletin, May, 1954.

Guilford, J.P., "Three Faces of Intellect". American Psychologist, 14: 469-479. 1959.

Gulliksen, Harold. Theory of Mental Tests. New York: Wiley and Sons, Inc., 1950.

Haggard, E.A., "Socialization, Personality, and Academic Achievement in Gifted Children". School Review. Vol. 65:4, 388-414, 1957.

Hebb, D.O., A Textbook of Psychology. Philadelphia: Saunders, 1958.

Hebb, D.O., The Organization of Behavior. New York: Wiley and Sons, 1949.

Hebb, D.O., "The Effects of Early Experience on Problem Solving at Maturity". Am. Psych. 2: 306-307, 1947.

Husén, T., "The Influence of Schooling Upon I.Q. Theoria. 17: 61-88, 1951.

Hymovitch, B., "The Effects of Experimental Variations on Problem Solving in The Rat". J. Com. Physiol. Psych. 45: 313-321, 1952.

Lorge, I., "Schooling Makes a Difference". Teach. Coll. Rec. 46: 483-492, 1945.

Lorge, I., and Thorndike, R.L., The Lorge-Thorndike Intelligace Tests, (Technical Manual). Boston: Houghton Mifflin Company, 1957.

Lorge, I., and Thorndike, R.L., The Lorge-Thorndike Intelligence Tests (General Manual). Boston: Houghton Mifflin Company, 1954.

Lynn, R., "Environmental Conditions Affecting Intelligence". Educ. Research, 1: 3, 1959.

MacArthur, R.S., "Assessing General Intellectual Ability for Adaptive Teaching Decisions". (Paper presented to Annual Meeting of the Canadian Psychological Association June, 1961).

MacArthur, R.S., "The Coloured Progressive Matrices as a Measure of General Intellectual Ability for Edmonton Grade III Boys". Alta. J. of Ed. Res. 6: 2, 67-75, 1960.

Martin, A., and Wiekers, J.E., "Raven's Coloured Progressive Matrices and the Wechsler Intelligence Scale for Children". J. Consult. Psych., 18: 143-144, 1954.

Meyer, W.J., and Bendig, A.W., "A Longitudinal Study of the Primary Mental Abilities Test". Journal of Ed. Psych. 52: 1, 50-66, 1961.

Otis Quick-Scoring Mental Ability Tests, Manual of Directions for Beta Test. New York: World Book Company, 1939.

Otis Quick-Scoring Mental Ability Tests: New Edition. Manual of Directions (Alpha Short Form). New York: World Book Company, 1936.

Piaget, J., The Psychology of Intelligence. London: Routledge and Kegan Paul, 1950.

Raven, J.C., Guide to Using the Coloured Progressive Matrices. London: Lewis and Company, Ltd., 1960.

Raven, J.C., Guide to Using the Progressive Matrices. London: Lewis and Company, 1954.

Riesen, A.H., "Arrested Vision". Scient. Am. 183: 16-19, 1950.

Safran Culturally Reduced Intelligence Test. Safran, C. Calgary Public School Board, 1960.

Thompson, W.R., "Early Environment -- Its Importance to Later Behavior" in Hoch, J.H. and Zubin, J. Psychopathology of Childhood, New York: Grune and Stratton, 1955.

Thompson, W.R., and Heron, W., "The Effects of Restricting Early Experiences on the Problem Solving Capacity of Dogs". Can. J. Psych., 8: 17-31, 1954.

Vernon, P.E., "The Classification of Abilities". Educational Research. 3: 2, 184-193, 1960.

Vernon, P.E., "A New Look at Intelligence Testing". Educational Research. 1: 1, 3-12, 1958.

Vernon, P.E., The Structure of Human Abilities. London: Mathuen, 1950.

Walker, H.M., and Lev, J., Statistical Inference. New York: Henry Holt, 1953.

Wechsler, David. The Measurement and Appraisal of Adult Intelligence. Baltimore: Williams and Wilkins Company, 1958.